

„TISSUE TENSIO METER“ UND „SPINELINER“ - UNTERSUCHUNG DER OBJEKTIVITÄT UND RELIABILITÄT AM MODELL

Dissertation
zur Erlangung des akademischen Grades

doctor medicinae (Dr. med.)

vorgelegt dem Rat der Medizinischen Fakultät
der Friedrich-Schiller-Universität Jena

von Anne-Maria Zschiegner
geboren am 07.10.1987 in Erfurt

Gutachter

1. Prof. Dr. med. Ulrich Smolenski (Jena)
2. PD Dr. med. Andreas Roth (Eisenberg)
3. PD Dr. med. habil. Johannes Buchmann (Rostock)

Tag der öffentlichen Verteidigung:

03.06.2013

Inhaltsverzeichnis

Abkürzungsverzeichnis.....	III
Abbildungsverzeichnis.....	IV
Tabellenverzeichnis.....	V
Diagrammverzeichnis.....	VI
Zusammenfassung.....	01
1 Einleitung.....	04
1.1 Manuelle Medizin.....	04
1.1.1 Funktionelle Störungen im Bewegungssystem.....	05
1.1.2 Diagnostik in der Manuellen Medizin.....	09
1.1.3 Entwicklung von Messgeräten in der Manualmedizin.....	12
1.2 Testgütekriterien.....	15
1.2.1 Objektivität.....	15
1.2.2 Reliabilität.....	16
1.2.3 Validität.....	17
1.3 Literaturanalyse.....	18
2 Ziele der Arbeit.....	22
2.1 Zielsetzung.....	22
2.2 Fragestellungen und Hypothesen.....	22
2.2.1 Fragestellungen.....	22
2.2.2 Hypothesen.....	22
3 Material und Methoden.....	24
3.1 Studiendesign.....	24
3.2 Aufbau und Funktionsweise der Geräte.....	24
3.2.1 „Tissue Tensiometer“.....	24
3.2.2 „Spineliner“.....	25
3.2.3 Übersicht.....	30
3.3 Gewebemodelle.....	31
3.4 Studienablauf.....	34
3.4.1 Messung mit dem „Tissue Tensiometer“.....	35
3.4.2 Messung mit dem „Spineliner“.....	37
3.5 Statistische Methoden.....	39
4 Ergebnisse.....	43
4.1 „Tissue Tensiometer“.....	43

4.1.1 Mittelwertvergleiche.....	43
4.1.2 Intrasession-Reliabilität.....	47
4.1.3 Objektivität.....	48
4.1.4 Test-Retest-Reliabilität.....	48
4.1.5 Interrater-Reliabilität.....	51
4.2 „Spineliner“	52
4.2.1 Mittelwertvergleiche.....	52
4.2.2 Intrasession-Reliabilität.....	57
4.2.3 Objektivität.....	58
4.2.4 Test-Retest-Reliabilität.....	59
4.2.5 Interrater-Reliabilität.....	62
5 Diskussion.....	65
5.1 Ergebnisdiskussion.....	65
5.1.1 Gütekriterien.....	65
5.1.2 „Tissue Tensiometer“	68
5.1.3 „Spineliner“	73
5.2 Methodendiskussion.....	78
5.2.1 Modelle.....	78
5.2.2 Ablauf und Aufbau der Studie.....	81
6 Fazit und Ausblick.....	85
7 Literaturverzeichnis.....	86
8 Anhang.....	95

Abkürzungsverzeichnis

BMI	Body Mass Index
BWS	Brustwirbelsäule
CE	<i>ursprünglich</i> Communauté Européenne
CROM	Cervical range of motion (Bewegungsumfang der HWS)
DGMM	Deutsche Gesellschaft für Manuelle Medizin
EMV	Elektromagnetische Verträglichkeit
FIMM	Fédération Internationale de Médecine Manuelle
HWS	Halswirbelsäule
ICC	Intraclass Correlation Coefficient
ISI	Institute for scientific information
LWS	Lendenwirbelsäule
M.	Musculus
MZP	Messzeitpunkt
N	Newton
n	Anzahl der Messpunkte
NTI	Neurogenic Technologies, Incorporated
PAS	Posterior-to-anterior spinal stiffness measuring system
RA	Rater A = Autor
RB	Rater B = Studienberater
SAM	Stiffness Assessing Machine
SD	standard deviation = Standardabweichung
SPSS	<i>ursprünglich</i> Statistical Package for the Social Sciences, Firma für Statistik- und Analyse-Software
TTC	TrainingTradingConsulting GmbH
TTM	Tissue Tensiometer

Abbildungsverzeichnis

Abbildung 1	Myofasziale Triggerpunkte und das „taut band“ (Davies 2006).....	05
Abbildung 2	Übertragungsschmerzmuster des M.latissimus dorsi.....	07
Abbildung 3	Kombinationen von Reliabilität und Validität (Harting et al. 2007).....	17
Abbildung 4	Das „Tissue Tensiometer“	24
Abbildung 5	Stempel des „TTM“	25
Abbildung 6	Schwingungskurve bei schwacher und starker Dämpfung (Grimsehl und Schallreuter 1991).....	26
Abbildung 7	Parameter der Schwingungskurve (Beyer et al. 2009).....	27
Abbildung 8	Aufbau des „Spineliners“ (TTC 2008).....	28
Abbildung 9	Handhabung des „Spineliners“	28
Abbildung 10	Auswahl der Behandlungsaufsätze (Rustler 2010).....	29
Abbildung 11	Aufbau eines Gewebemodells.....	31
Abbildung 12	Deckmatte des Modell 3A mit Markierungen.....	32
Abbildung 13	Zwei Modelle, Schichten dargestellt.....	34
Abbildung 14	Studienablauf.....	35
Abbildung 15	Untersuchungsaufbau „TTM“	36
Abbildung 16	Bildschirmausgabe nach einer Messung mit dem „TTM“	36
Abbildung 17	Untersuchungsaufbau „Spineliner“	37
Abbildung 18	Ausdruck einer „Spineliner“-Messung.....	38
Abbildung 19	Formel zur Berechnung der Interrater-ICCs für zufällige Rater (Eliasziw et al.1994).....	41

Tabellenverzeichnis

Tabelle 1	Übersicht der Eigenschaften des „TTM“ und des „Spineliners“.....	30
Tabelle 2	Modelle mit Zuordnung der Schichten.....	33
Tabelle 3	Interpretation der ICC-Werte im Hinblick auf die Reliabilität (Bortz et al. 2008).....	39
Tabelle 4	Mittelwerte des „TTM“ pro Rater über alle MZP in N/m ² (n=60 pro Modell, n=1200 über alle Modelle).....	43
Tabelle 5	Gruppenunterschied „TTM“ pro Rater über alle MZP (pro Rater n=1200).....	46
Tabelle 6	ICC- und p-Werte der Intrasection-Reliabilität des „TTM“ von Rater A über alle und zwischen zwei Messungen (Overall n=30, zwischen zwei Messungen n=20).....	47
Tabelle 7	Test-Retest-Reliabilität des „TTM“ pro Rater über alle und zwischen zwei MZP (Overall _{Einzel} n=1200, zwischen zwei MZP n=480).....	49
Tabelle 8	Interrater-Reliabilität des „TTM“ zwischen den Ratern über alle und pro MZP (Overall n=2400, pro MZP n=480).....	51
Tabelle 9	Mittelwerte des „Spineliners“ pro Rater über alle MZP (n=60 pro Modell, n=1200 über alle Modelle).....	52
Tabelle 10	Gruppenunterschied „Spineliner“ pro Rater über alle MZP (pro Rater und Parameter n=1200).....	55
Tabelle 11	Intrasection-Reliabilität des „Spineliners“ von Rater A über alle und zwischen zwei Messungen (Overall n=30, zwischen zwei Messungen n=20).....	57
Tabelle 12	Objektivität (Korrelation r) aller vier Parameter des „Spineliners“ zwischen den Ratern pro MZP (n=480).....	58
Tabelle 13	Test-Retest-Reliabilität (ICC) des „Spineliners“ pro Rater über alle und zwischen zwei MZP (Overall _{Einzel} n=1200, zwischen zwei MZP n=480)	59
Tabelle 14	Interrater-Reliabilität des „Spineliners“ zwischen den Ratern über alle und pro MZP (Overall n=2400, pro MZP n=480).....	63

Diagrammverzeichnis

Diagramm 1	Mittelwertvergleich der Modelle mit „TTM“ zwischen den Ratern über alle MZP (n=60; entspricht pro Rater 5 Messungen bei 12 Punkten je Modell).....	44
Diagramm 2	Minimum, Mittelwert und Maximum des „TTM“ pro Rater und pro MZP (n=240).....	45
Diagramm 3	Verteilungskurve der Messwerte des „TTM“ pro Rater zu MZP 1 (pro Diagramm n=240).....	47
Diagramm 4	Objektivität (Korrelation r) des „TTM“ zwischen den Ratern pro MZP (jeweils n=480).....	48
Diagramm 5	Test-Retest-Reliabilität des „TTM“ pro Rater über alle und zwischen zwei MZP (Overall _{Einzel} n=1200, zwischen zwei MZP n=240).....	50
Diagramm 6	Mittelwerte des Widerstands der einzelnen Modelle pro Rater über alle MZP (n=60).....	53
Diagramm 7	Minimal-, Mittel- und Maximalwerte der Mobilität pro Rater und pro MZP (n=240).....	54
Diagramm 8	Minimal-, Mittel- und Maximalwerte der Frequenz pro Rater und pro MZP (n=240).....	54
Diagramm 9	Beispiel Häufigkeitsverteilung der Mobilität pro Rater zu MZP 1 (pro Diagramm n=1200).....	56
Diagramm 10	Beispiel Häufigkeitsverteilung der Frequenz von Rater A zu MZP 3 und 4 (pro Diagramm n=1200).....	56
Diagramm 11	Intrasession-Reliabilität des „Spineliners“ von Rater A über alle und zwischen zwei Messungen (Overall n=30, zwischen zwei Messungen n=20).....	57
Diagramm 12	Objektivität (Korrelation r) aller vier Parameter des „Spineliners“ zwischen den Ratern pro MZP (n=480).....	58
Diagramm 13	Test-Retest-Reliabilität des „Spineliners“ 1 pro Rater zwischen zwei MZP (n=480).....	60
Diagramm 14	Test-Retest-Reliabilität des „Spineliners“ 2 pro Rater zwischen zwei MZP (n=480).....	61
Diagramm 15	Test-Retest-Reliabilität des „Spineliners“ 3 pro Rater über alle MZP (n=1200).....	62

Diagramm 16	Interrater-Reliabilität des „Spineliners“ zwischen den Ratern über alle und pro MZP (Overall n=2400, pro MZP n=480).....	63
-------------	--	----

Zusammenfassung

Einleitung: Die Untersuchungsmethoden der Manuellen Medizin unterliegen einer hohen Subjektivität. Um diese einzudämmen, werden zunehmend technische Apparate zur Unterstützung und Objektivierung der Diagnostik eingesetzt. Es muss zunächst geprüft werden, welche Geräte eine diagnostische Bereicherung darstellen. Zu diesem Zweck werden die Gütekriterien Objektivität, Reliabilität und Validität in kontrollierten Studien getestet.

Fragestellung: Diese Studie widmet sich der Überprüfung von Objektivität und Reliabilität an dem „Tissue Tensiometer“ und dem „Spineliner“ mithilfe von Modellen. Zielgröße ist eine Reliabilität von ICC $\geq 0,5$.

Material und Methoden: Es wurden zwanzig Holzkästen entwickelt, in die verschiedene Matten und Schaumstoffe zur Gewebeimitation eingelegt wurden. Zehn davon waren zusätzlich mit einem Holzbrett versehen, um eine Verhärtung darzustellen. Auf diesen Modellen wurden 240 Messpunkte markiert. Eine mittige Position der Modelle auf einer Untersuchungsfläche war Bestandteil des festgelegten Messaufbaus. Die Messungen wurden mit den Geräten „Tissue Tensiometer“ und „Spineliner“ realisiert. Der Messablauf konnte in einer Trainingsphase erlernt und standardisiert werden. Zunächst erfolgte eine Intrasession-Messung an zehn Messpunkten durch einen Rater. Es folgten für beide Geräte fünf Messtermine mit einem einwöchigen Abstand. An diesen Terminen wurde durch zwei Rater gemessen. Aus den Daten ergibt sich die Objektivität, Test-Retest-Reliabilität und Interrater-Reliabilität.

Ergebnisse: „TTM“: Intrasession-Reliabilität zufriedenstellend (ICC=0,76), Objektivität ausreichend bis zufriedenstellend zu MZP 4 ($r=0,71$) und 5 ($r=0,68$). Test-Retest-Reliabilität (Overall) nicht ausreichend, Interrater-Reliabilität (Overall) nicht ausreichend. „Spineliner“: Intrasession-Reliabilität ausreichend bis zufriedenstellend (ICC=0,51-0,76), Objektivität zu MZP 3 für alle Parameter ausreichend bis zufriedenstellend ($r=0,61$ -0,87), für Frequenz außerdem zu MZP 2-5 ($r=0,55$ -0,82), Test-Retest-Reliabilität (Overall) nicht ausreichend, Interrater-Reliabilität (Overall) ausreichend für Widerstand, Frequenz und Mobilität.

Unterschiede zwischen den Modellgruppen konnten durch signifikante Mittelwertunterschiede nachgewiesen werden.

Diskussion: Beide Geräte wiesen starke Schwankungen der Reliabilitäten zwischen den MZP auf. Die Interrater-Reliabilität ist höher als die Test-Retest- bzw. Intrarater-

Reliabilität. Es wird der Einfluss von elektromechanischen Störfaktoren diskutiert. Der Versuchsaufbau und -ablauf wurde standardisiert eingehalten. Die Modelle scheinen dem Anspruch der Reliabilitätsmessung gerecht zu sein. Modellcharakteristika wurden von beiden Geräten erkannt, jedoch wären hier einheitlichere Modellgruppen von Vorteil gewesen. Eine kombinierte Betrachtung der Werte mit den Messkurven kann zu einer erhöhten Aussagekraft der Messung beitragen.

Schlussfolgerung: Eine rein auf die Messwerte bezogene Diagnosestellung ist nicht zu empfehlen. Vielmehr sollten die Messwerte individuell und im Verlauf betrachtet werden. Die Geräte sind als Diagnose- und Dokumentationsunterstützung geeignet, können und sollen jedoch den Arzt und seine persönliche Erfahrung am Patienten nicht ersetzen.

Summary

Introduction: The methods of manual medicine are characterized by high subjectivity. To mitigate this, the use of technical equipment to support and objectify the diagnostics increased. It has to be analyzed, which devices represent a diagnostic enrichment. For this purpose, the quality criteria objectivity, reliability and validity are tested in controlled studies.

Objectives: This study is devoted to investigate the objectivity and reliability of the "Tissue Tensiometer" and the "Spineliner" using models. The aim is a reliability of ICC ≥ 0.5 .

Methods: Twenty wooden boxes were developed, in which various mats and foams were placed for tissue imitation. Ten of them were additionally provided with a wooden board to represent a hardening. On these models, 240 measuring points were marked. The models were positioned in the middle of an examination table. Measurements were undertaken with the devices "Tissue Tensiometer" and "Spineliner". In a training period the measurement procedure was practised and standardized. First, an intrasession measurement took place at ten measuring points by one rater. For both devices, five measuring dates followed with a one-week interval. On these dates two raters were involved. The objectivity, test-retest reliability and interrater reliability result from this data.

Results: "TTM": Intrasession reliability satisfactorily (ICC=0.76), objectivity sufficient to satisfactorily at MZP (time of measurement) 4 ($r=0.71$) and 5 ($r=0.68$). Test-retest reliability (overall) is not sufficient, interrater reliability (overall) is not sufficient.

"Spineliner": Intrasection reliability sufficient to satisfactory (ICC=0.51 to 0.76), objectivity at MZP 3 for all parameters sufficient to satisfactory (r=0.61 to 0.87), for frequency additionally at MZP 2-5 (r=0.55 to 0.82), test-retest reliability (overall) is not sufficient, interrater reliability (overall) is sufficient for the parameters resistance, frequency and mobility. Differences between the model groups could be demonstrated by significant differences in means.

Discussion: Both devices showed large variations in the reliabilities between the MZP. The interrater reliability is higher than the test-retest or intrarater reliability. An influence of electro-mechanical disturbances is discussed. The experimental setup and procedure was standardized. The models seem to be suited for measures of reliability. Model characteristics were detected by both devices, but consistent groups of models would have been preferable. A combined analysis of the data with the measuring curves can contribute to increased reliability of the measurement.

Conclusion: A purely data-based diagnosis is not recommended. Rather, the data should be considered individually and through a chronological perspective. The devices are available as a support for diagnostics and documentation. They cannot and should not replace the doctor and his personal experience on the patient.

1 Einleitung

1.1 Manuelle Medizin

Seit Menschengedenken wird bei Krankheiten oder Verletzungen manuell untersucht und, wenn möglich, durch einfache Handgriffe therapiert (Beyer 2008). Die wahrscheinlich älteste Abbildung zur Thematik stammt aus Indien und der Zeit zwischen 3500 und 1800 v. Chr. Sie zeigt die Streckung der Wirbelsäule durch Fixation der Füße und Anheben des restlichen Körpers (Kumar 1996). Schriftliche Berichte über verschiedene Grifftechniken wurden bereits von den Chaldäern aus Babylonien in den Jahren 1000 bis 600 v. Chr. erstellt. Hippokrates (um 400 v. Chr.) und Galen (um 150 n. Chr.) vertieften und verbreiteten das Wissen um die heilende Wirkung von Traktion (Zug) an Gelenken (Cramer 1990). Für lange Zeit wurden die Behandlungstechniken des „Knochensetzens“ in der Volksmedizin weitergegeben. Erst um das Jahr 1900 erfuhr die Manualmedizin wissenschaftlichen Aufschwung durch die Gründung von Schulen für Osteopathie durch Andrew Taylor Still und für Chiropraktik durch Daniel David Palmer (Greenman 1998, Deutsche Gesellschaft für Manuelle Medizin 2011). Heutzutage lässt sich Manuelle Medizin folgendermaßen definieren:

*Manuelle Medizin bezeichnet die Lehre der
Erkennung und Behandlung von Funktionsstörungen an den
Haltungs- und Bewegungsorganen mithilfe spezieller Handgriffe (Krämer 2006).*

Die Hand des Untersuchers nimmt in diesem Fachgebiet bezüglich Diagnostik und Behandlung einen großen Stellenwert ein, was sich auch in der Wortherkunft widerspiegelt. „Manuell“ leitet sich vom lateinischen „manus“ für Hand ab. Die Vorsilbe „chiro-“ stammt von dem griechischen Wort „χείρα“ [çira], übersetzt mit „Hand“ (PONS 2008). In Deutschland wird „Chirotherapie“ als Synonym für „Manuelle Medizin“ verwendet, nicht zu verwechseln mit „Chiropraktik“ und „Osteopathie“. Letztere Praktiken werden hierzulande von Nicht-Ärzten ausgeübt. Sie entstanden in den USA, während sich in Europa die Manuelle Medizin etablierte (Baumgartner et al. 1993). Die drei Disziplinen zeigen große Übereinstimmungen in Diagnostik und Therapie, die Grenzen zwischen den Fachgebieten verlaufen zum Teil nicht sehr trennscharf (Dvorak et al. 2001). An einer stärkeren Verschmelzung von chirotherapeutischen und osteopathischen Methoden und der Einbindung in die ärztliche Aus-

und Weiterbildung in Deutschland besteht zunehmend Interesse (Buchmann 2002). Zur Diagnostik werden verschiedene manuelle Tests angewandt, um Funktionsstörungen an Muskulatur, Wirbelsäule oder Extremitätengelenken festzustellen. Diese werden im Rahmen der Manuellen Therapie durch besondere Techniken behandelt, bspw. durch Mobilisation oder Manipulation. Erhöhte Gewebespannung in Form von muskulären Triggerpunkten oder eine reversible segmentale Dysfunktion der Wirbelsäule gehören zu den häufigsten Indikationen zur Manuellen Therapie (Heimann 2008). Weitere Symptome können durch den Manualmediziner abgeklärt werden: Schwindel, kranio-mandibuläre Dysfunktion, Tinnitus, Kopfschmerz, Stereotypstörungen und muskuläre Dysbalancen.

1.1.1 Funktionelle Störungen im Bewegungssystem

Anthropologisch betrachtet, ist der Mensch nicht für lang andauernde sitzende Tätigkeiten oder einseitige Belastungen geschaffen. Trotzdem erfordert das moderne Berufsleben immer weniger Bewegung und immer mehr Schreibtischarbeit. Chronisch falsche Körperhaltungen, bspw. ständiges Anheben des Schultergürtels beim Bedienen der Computertastatur oder ein permanent gekrümmter Rücken, führen zu funktionellen Störungen im Bewegungssystem (Muscolino 2009). Diese sind reversibel, können aber zu strukturellen Schädigungen beitragen.

Durch Fehlbeanspruchung eines Muskels entstehen myofasziale Triggerpunkte. Dies sind tastbare Befunde im Muskel, die sehr schmerzhaft auf Druck reagieren. Simons und Travell 1999 beschreiben außerdem ein sogenanntes „taut band“ (Hartspannstrang), welches vom Triggerpunkt aus in beide Richtungen der Verankerung des Muskels zieht und die Festigkeit einer Sehne haben kann (siehe Abbildung 1).

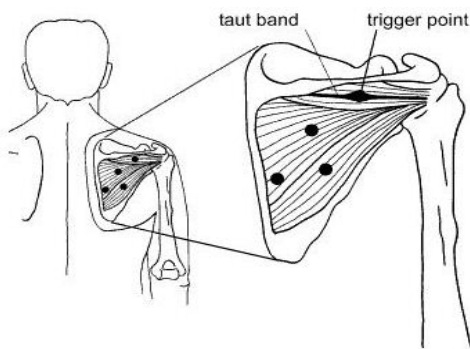


Abbildung 1: Myofasziale Triggerpunkte und das „taut band“ (Davies 2006)

Dadurch wird das passive Bewegungsausmaß eines Muskels gehemmt, da er sich nicht mehr ausreichend verlängern kann. Ein Triggerpunkt mit taut band ist ein chronischer Kontrakturstrang innerhalb des Muskels, nicht zu verwechseln mit einem plötzlich auftretenden und schnell zu beseitigenden Muskelkrampf (Davies 2006).

Es werden aktive von latenten Triggerpunkten unterschieden. Aktive Triggerpunkte können bereits in Ruhe oder bei physiologischer Muskelbelastung Schmerzen auslösen. Sie sind im körperlich aktiven Lebensalter am wahrscheinlichsten. Typisch ist das sogenannte „jump sign“ bei Reizung des Triggerpunktes: eine unwillkürliche Zuckungs- oder Abwehrreaktion des Patienten (Simons und Travell 1999).

Mit Abnahme der Leistungsfähigkeit im fortgeschrittenen Alter steht die Bewegungseinschränkung durch latente Triggerpunkte, die erst bei deutlicher Reizung durch Nadeln oder starken Druck Schmerzen ausstrahlen, im Vordergrund. Deren Diagnose ist besonders schwierig (Eder und Tilscher 1990; Simons und Travell 1999).

Knapp 50 Prozent des menschlichen Körpergewichts entfallen auf die Skelettmuskulatur und jeder dieser rund 400 Muskeln kann Triggerpunkte entwickeln (Simons und Travell 1999). Die Bedeutung der Triggerpunkte sollte nicht unterschätzt werden, da sie (auch in entfernten Körperregionen) Schmerzen, motorische Dysfunktion und autonome Phänomene (wie Vasokonstriktion oder Hypersekretion) hervorrufen können. Dieses Phänomen wird als „referred pain“ (deutsch: Übertragungsschmerz) bezeichnet. Für die typischen Triggerpunktlokalisationen wurden bestimmte Übertragungsschmerzzone beschrieben, die bei nahezu allen Patienten vorhanden sind (Hauptschmerzzone). Die bei einigen Patienten darüber hinausgehende Schmerzausstrahlung wird als Nebenschmerzzone bezeichnet (Simons et al. 2002).

Myofasziale Triggerpunkte sind schmerzhaft lokale Verhärtungen innerhalb eines Hartspannstrangs („taut band“) im Muskelgewebe, die auch in entfernten Körperregionen Übertragungsschmerz („referred pain“) und vegetative Symptome auslösen können (Simons und Travell 1999).

Beispielhaft zeigt Abbildung 2 mögliche Übertragungsschmerzmuster von Triggerpunkten im M.latissimus dorsi. Die Triggerpunkte sind mit einem x markiert, die Hauptschmerzzone ist in einem flächigen und die Nebenschmerzzone in einem punktierten Rot dargestellt (Simons und Travell 2004).

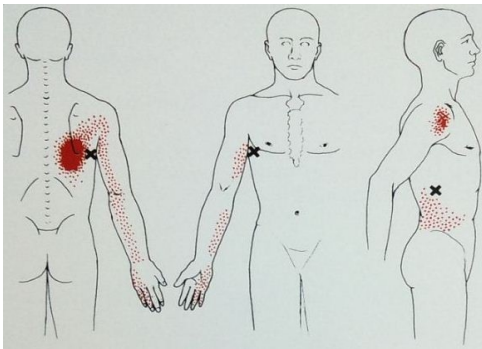


Abbildung 2: Übertragungsschmerzmuster des M.latissimus dorsi

Eine Differentialdiagnose für Schmerzen und Bewegungseinschränkungen ist die reversible segmentale Dysfunktion. Reversibel, da keine strukturellen Schäden nachweisbar sind und sich die Dysfunktion somit ad integrum zurückbilden kann. „Segmental“ bezieht sich auf die Wirbelsäule - die Dysfunktion tritt innerhalb verschiedener Segmente auf.

Als Ursache versteht man heutzutage eine Kombination an Störungen im periartikulären Bereich, auch Arthron genannt, der sich aus drei Systemen zusammensetzt:

- dem artikulären System mit Gelenkkapsel und Bändern (passiv bewegt),
- dem myofaszialen System (aktive Bewegung) und
- dem neuroreflektorischen System (Steuerung der Bewegung) (Kayser 2010).

Der alte Begriff der „Blockierung“ wich zunehmend dem der „reversiblen segmentalen Dysfunktion“, um die Assoziation eines rein mechanischen Problems zu verhindern. Frühere Theorien gingen von einem alleinigen Ursprung innerhalb des Gelenks aus, bspw. durch Einklemmen eines seiner Bestandteile.

Erst später wurde der Rolle der Neurophysiologie größere Beachtung geschenkt: ist das Zusammenspiel zwischen Halte- und Bewegungsmotorik gestört, kommt es zur segmentalen Dysfunktion. Neben Bewegungsmangel bzw. Fehlbelastung können auch Traumata und Stoffwechselstörungen bis hin zu psychischen Belastungen oder

viszeralen Erkrankungen ursächlich sein. Die komplexe neuronale Verschaltung zwischen Muskelapparat, Gehirn, Haut und inneren Organen bewirkt eine ständige Beeinflussung des Muskeltonus durch äußere Reize (Heimann 2008; Speckmann und Wittkowski 1997).

Die gelenkumgebende Muskulatur kann durch Tonuserhöhung zu einer Dysfunktion beitragen. Das bestätigen auch Goldmann et al. 1997: in ihrer Studie an 58 Patienten mit segmentaler Dysfunktion konnten diese bei acht Patienten unter Muskelrelaxation nicht mehr nachgewiesen werden, bei sechs davon trat sie nach Narkose erneut auf. Das spricht für ein multifaktorielles Geschehen. Buchmann et al. 2005 behandelten 26 Patienten mithilfe von zwei manualmedizinischen Techniken jeweils vor, während und nach einer Narkose. Es zeigte sich, dass postisometrische Relaxation vor allem die Muskeln anspricht, während spinale Manipulation zusätzlich auf Gelenkkapsel, Bänder und Faszien zu wirken scheint.

Die Gesamtheit dieser Pathologien (artikulär, myofaszial und neuroreflektorisch) kann zu Schmerzen und Hypomobilität führen. Dabei ist das Gelenkspiel beeinträchtigt, aber nie ganz aufgehoben. Jedes Gelenk ist durch ein Bewegungsausmaß gekennzeichnet, welches sich aus der aktiven und passiven Beweglichkeit zusammensetzt und normalerweise erst durch die anatomische Barriere begrenzt wird. Bei einer segmentalen Dysfunktion ist das Bewegungsausmaß eingeschränkt, man spricht von einer pathologischen Barriere (Kayser 2010).

In der Wirbelsäule reagiert das Nachbarsegment eines funktionsgestörten Wirbelgelenks zunächst mit Hypermobilität. Wird das Problem nicht beseitigt, generiert sich im Gewebe durch den erhöhten mechanischen Reiz eine Entzündung mit gleichzeitiger Bildung von Osteophyten. Es entsteht eine tellerrandartige Verbreiterung der Auflagefläche der Wirbelkörper. Ist ein Wirbel derartig verändert, können durch die abnormale Belastung weitere Segmente geschädigt werden, es kommt zu Abnutzung und Arthrose (Lewit 1997). Daher ist es wichtig, frühzeitig mithilfe manueller Therapie einzugreifen. Dadurch können irreversible strukturelle Störungen verhindert werden.

Für beide Formen der funktionellen Störungen gilt: meist steht für den Patienten der Schmerz im Vordergrund. Er kann in der betroffenen Region lokalisiert sein oder in eine andere Region ausstrahlen. Das Leitsymptom Schmerz ist der Auslöser für den

Arztbesuch und der Beginn des diagnostischen und therapeutischen Regimes.

1.1.2 Diagnostik in der Manuellen Medizin

Betritt der Patient das Untersuchungszimmer, liefert er dem Arzt durch sein Gangbild, seine Körperhaltung und seine Konstitution bereits erste Hinweise auf eventuell vorhandene Funktionsstörungen.

Im Anschluss erfolgt, wie in jeder klinischen Disziplin, eine ausführliche Beschwerdeanamnese. Zunächst ist zu erfragen, welche Beschwerden bereits wie lange andauern, ob es einen akuten Auslöser gab und was bisher dagegen unternommen wurde. Als Hauptproblem werden in den meisten Fällen Schmerzen angegeben. Bewegungseinschränkungen fallen teilweise erst auf, wenn sie den Alltag des Patienten beeinträchtigen. Zu Hinterfragen sind körperliche Aktivitäten des Patienten, eine eventuelle Schmerzausstrahlung, Beeinträchtigungen der Sensibilität oder Motorik, Nebenerkrankungen und Medikation. Denn in manchen Fällen stehen hinter Rückenschmerzen Bandscheibenvorfälle, osteoporotisch bedingte Veränderungen oder mögliche internistische Krankheitsbilder. Die Häufigkeit dieser spezifischen Schmerzen beträgt weniger als 20% aller Rückenschmerzen (Diemer und Sutor 2007). Die unspezifischen Rückenschmerzen machen den Großteil aus und stehen im Fokus der Manualmedizin.

Die körperliche Untersuchung beginnt mit der gründlichen Inspektion des Patienten zum Erkennen eventuell vorhandener Entzündungszeichen, Muskelatrophien oder Fehlstellungen (Heimann 2008).

Orientierend wird das Bewegungsausmaß der HWS, BWS und LWS hinsichtlich des Bewegungsumfangs, der Seitengleichheit und eines eventuellen Schmerzauftritts überprüft. Sogar das Ende der Bewegung kann durch einen harten oder weichen Anschlag diagnostisch richtungsweisend sein. Bei der Wirbelsäulenuntersuchung gibt es sechs Bewegungsrichtungen: Anteflexion und Retroflexion sowie Lateroflexion und Rotation je nach links und rechts.

Die Prüfung erfolgt aktiv mit und ohne Widerstand sowie passiv. Bei der aktiven Prüfung wird der Patient aufgefordert, die Bewegungen selbst in dem für ihn möglichen Ausmaß durchzuführen. Zur Widerstandsprüfung leistet der Untersucher eine Gegenkraft und testet damit die Stärke des Muskel- und Sehnensystems. Bei der passiven Prüfung bewegt der Untersucher den Patienten ohne dessen Mithilfe.

Durch diese drei Verfahren kann der Manualmediziner Rückschlüsse auf die Ursache einer Pathologie ziehen: arthrogen, myogen oder neurologisch (Eder und Tilscher 1990). Arthrogene Störungen äußern sich sowohl aktiv als auch passiv, da die Bewegung an einer bestimmten Stelle durch anatomische Voraussetzungen blockiert wird oder Schmerzen verursacht. Ist die aktive Bewegung im Vergleich zur passiven Bewegung deutlich eingeschränkt, könnte eine neurologische Schädigung dahinter stecken, die eine regelrechte Überleitung des Bewegungsimpulses von Gehirn zum Muskel verhindert. Ein erhöhter Muskeltonus auch während der passiven Bewegung kann ebenfalls neurologisch begründet sein, oder für eine myogene Ursache im Sinne von Triggerpunkten sprechen. Zur genauen Differenzierung sind alle Informationen aus Anamnese und klinischer Untersuchung einzubeziehen.

Falls bei der orientierenden Untersuchung Beeinträchtigungen festgestellt wurden, findet eine regionale Untersuchung durch Palpation der einzelnen Bewegungs-segmente oder Extremitätengelenke und ihrer Umgebung statt. Palpatorisch ist eine segmentale Dysfunktion durch eine Spannungserhöhung des Gewebes zu erkennen, die auf eine Hypomobilität zurückzuführen ist (Kayser 2010). Ebenso sind myofasziale Triggerpunkte anhand ihrer erhöhten Festigkeit gut palpierbar. Neben der Strukturdiagnose kann man auch Schmerzen, thermische und mechanische Qualitäten erfühlen und mithilfe der Anamnese zu einer sinnvollen Diagnose gelangen. Die Basis für eine solche Beurteilung sind fundierte anatomische Kenntnisse kombiniert mit langjähriger Erfahrung hinsichtlich normaler und pathologischer Tastbefunde.

Es wird deutlich, dass die Diagnostik in der Manuellen Medizin sehr durch die körperliche Untersuchung statt durch bildgebende Verfahren oder Laborparameter geprägt ist. Letztere werden meist zu differentialdiagnostischen Zwecken (Entzündungen, Tumore, Osteoporose, Arthrose usw.) angefordert.

Bei funktionellen Störungen sind die Befunde nicht metrisch festzulegen und somit ist ein nachfolgender Untersucher auf das Deskriptionstalent des Vorhergehenden angewiesen (Psczolla 1997). Dementsprechend ist die Dokumentation von Messgrößen problematisch, da viele manualmedizinische Befunde subjektiven Einschätzungen und Wertungen des Arztes unterliegen. Diese Subjektivität kommt durch die unterschiedliche Ausbildung und Erfahrung jedes Arztes zustande.

In Deutschland existiert eine Dachgesellschaft für Manualmediziner: die Deutsche Gesellschaft für Manuelle Medizin (DGMM). Die Zielsetzung der DGMM ist die „wissenschaftliche Erforschung und Weiterentwicklung der Manuellen Medizin sowie ihre Integration in die medizinische Versorgung“. Der DGMM gehören drei Ärzteseminare an, die mit verschiedenen Ausbildungscurricula zur Erlangung der Zusatzbezeichnung Manuelle Medizin/ Chirotherapie führen. Voraussetzung ist eine Facharztanerkennung, unabhängig von der Fachrichtung (Deutsche Gesellschaft für Manuelle Medizin 2011). Somit hat jeder Teilnehmer unterschiedliche Vorkenntnisse. Die manualmedizinischen Schulen unterscheiden sich, sowohl in Deutschland als auch weltweit, in den gelehrt diagnostischen und therapeutischen Verfahren. Daher hat die Fédération Internationale de Médecine Manuelle (FIMM) Protokolle zur Durchführung von Reliabilitäts- und Validitätsstudien bezüglich manualmedizinischer Tests herausgegeben, um geeignete Tests herauszufiltern und damit die Lehre zu standardisieren (Patijn 2002).

Eine multizentrische Reliabilitätsstudie der DGMM wurde von Conradi et al. 2003 durchgeführt - es testeten jeweils vier Untersucher der drei deutschen Manualschoolen. An Low-back-pain-Patienten wurden elf Untersuchungstechniken angewendet und mithilfe eines vorher erarbeiteten Untersuchungsbogens dokumentiert. Allein bei der Auswahl der Techniken bestand große Uneinigkeit, die erst über mehrere Sitzungen hinweg (sog. Delphi-Verfahren) beseitigt werden konnte. Daran ist die Vielfalt der Untersuchungsmöglichkeiten und die Heterogenität der verschiedenen Ausbildungen zu erkennen. Diese spiegelte sich zudem in den Ergebnissen wider: viele Tests zeigten keine, wenige moderate und sehr wenige gute Übereinstimmungen (Conradi et al. 2003).

Metaanalysen widmeten sich der Frage nach Reproduzierbarkeit von Palpationstests zwischen verschiedenen Untersuchern. Die Qualität der Studien war sehr unterschiedlich: Conradi und Smolenski 2005 fanden keine Veröffentlichung auf dem Niveau der höchsten Evidenzstufe. Andere Autoren überprüften die Palpation segmentaler Beweglichkeit und die statische Wirbelsäulenuntersuchung (Haneline et al. 2008, Haneline und Young 2009). Die ermittelten Ergebnisse ließen keine allgemeingültige Aussage zur Überlegenheit eines bestimmten Palpationstests zu. Stochkendahl et al. 2007 betonen zudem, dass Tests, die in Kombination mit anderen Tests durchgeführt wurden, auch nur in exakt dieser Kombination die festgestellte Reliabilität besitzen. Zudem wird konstatiert, dass die Ergebnisse orientierender

Untersuchungen reliabler sind als Palpationstests spezieller Segmente. Weiterhin liefert die Untersuchung am stehenden Patienten schlechtere Ergebnisse als im Sitzen oder Liegen (Stochkendahl et al. 2007). In allen Reviews sowie einem Großteil der überprüften Studien fällt auf, dass die Übereinstimmung bei Schmerzprovokationstests im Allgemeinen größer ist als bei Mobilitäts- oder Funktionstests (Najm et al. 2003, Conradi und Smolenski 2005, Stochkendahl et al. 2007, Haneline und Young 2009).

Deutlich wird auch die Überlegenheit der Intrarater-Reliabilität gegenüber der Interrater-Reliabilität. Diese Tendenz ist zu erwarten, da unterschiedliche Untersucher auch unterschiedliche Voraussetzungen mitbringen. Ein einzelner Untersucher kann seine Ergebnisse zuverlässiger reproduzieren. Allerdings ist im klinischen Alltag eine Übertragbarkeit auf andere Untersucher notwendig.

1.1.3 Entwicklung von Messgeräten in der Manualmedizin

Im vorhergehenden Abschnitt wurde die Subjektivität und Vielfalt der diagnostischen Möglichkeiten in der Manualmedizin beschrieben. Die Interrater-Reliabilität von Palpationsbefunden ist oftmals unzureichend. Daraus ergibt sich, dass die Reproduzierbarkeit von Befunden erschwert ist, besonders bei einem Therapeutenwechsel. In der Ausbildung sollte mehr Wert auf die Reproduzierbarkeit und Validität von manuellen Untersuchungen gelegt werden, fordert Ammer 2009. Die Vereinheitlichung der Weiterbildung und manualmedizinischen Untersuchung wird bei der Zielsetzung der DGMM an erster Stelle genannt. Es gab bereits Bemühungen zur Senkung der Subjektivität, bspw. das Angleichen der Curricula der DGMM und Entwicklung von Ausbildungsstandards (Deutsche Gesellschaft für Manuelle Medizin 2011). Studien widmeten sich der Frage nach Richtlinien für Palpationstests, z.B. für manuelle Muskeltests (Schmitt und Cuthbert 2009). Andere entwarfen ein Palpationstrainingsgerät, um die Untersuchenden auf ein gemeinsames Level zu bringen (Anders et al. 2010). Doch es besteht weiterhin Forschungsbedarf auf diesem Gebiet, da die Reliabilität der Befundung in vielen Studien nicht zufriedenstellend ausfiel (Conradi und Smolenski 2005, Stochkendahl et al. 2007, Haneline und Young 2009).

Um die Diagnostik unabhängiger von einem speziellen Untersucher zu gestalten, liegt der Gedanke nahe, ein technisches Hilfsmittel zu nutzen. Somit werden

Einflussfaktoren wie Erfahrung, Tagesform und Talent eines Untersuchers minimiert. Ganz ausschalten kann man sie nicht, da auch die Handhabung eines Geräts von Untersuchern abhängig ist. Nachfolgend soll eine Auswahl an Gerätestudien der letzten Jahre dargestellt werden, um die Tendenz hinsichtlich apparativ gestützter Diagnostik im Bereich des menschlichen Bewegungssystems zu veranschaulichen.

Für die Diagnostik der Muskulatur wurden bereits verschiedene Geräte unter dem Namen „Myotonometer“ entworfen. Ältere Modelle nutzten eine invasive Technik zur Messung mithilfe von Nadeln innerhalb des Muskels (Steinlechner 1938), während neuere Entwicklungen eine non-invasive Technik bevorzugen. Vorteil ist eine schmerzfreie oder schmerzarme Untersuchung des Patienten. Das Patent für die Bezeichnung „Myotonometer“ besitzt mittlerweile die Firma „Neurogenic Technologies“ aus Montana (USA). Ein „Stempel“ wird von außen über einen Muskel gesetzt und mit einer definierten Kraft eingedrückt. Durch Bestimmung der Eindringtiefe des Stempels ins Gewebe sollen Rückschlüsse auf den Tonus des Muskels getroffen werden können (NTI 2011). Das „Myotonometer“ kann für Untersuchungen des Muskeltonus als reliabel betrachtet werden (Leonard 2003).

Ein leicht zu handhabendes aber wirkungsvolles Werkzeug zur Gelenkuntersuchung ist das Goniometer, ein Winkelmesser. Nützlich ist es zur Bestimmung von Bewegungsausmaßen rund um die Wirbelsäule und peripheren Gelenke. Spezielle Formen wurden entwickelt, wie das CROM- oder das Spin-T-Goniometer, welche an die menschliche Kopfform angepasst wurden. Gemessen an radiologischen Untersuchungen erreichte das CROM-Goniometer z.B. zur Messung der Flexion im HWS-Bereich eine gute Validität (Tousignant 2002). Auch das Spin-T-Goniometer ist im Vergleich zum Elektrogoniometer oder zu Ultraschall- und elektromagnetisch gestützten Verfahren wie dem „MotionStar“ eine preiswerte und valide Alternative (Agarwal 2005). Die genaue Angabe von Bewegungsausmaßen in Grad ist notwendig und seit einiger Zeit im Klinikalltag üblich. Nur so kann eine Verlaufsbeobachtung auch arztübergreifend stattfinden.

Im Zuge des medizinischen Fortschritts wurden Geräte entworfen, welche die Abschätzung von Bewegungsausmaßen noch komfortabler gestalten. Die „Spinal Mouse“ zum Beispiel muss für die Messung der Wirbelsäulenbeweglichkeit einfach über den Rücken gerollt werden (Kellis 2008). Das „Zebris“-Ultraschall-Messsystem

funktioniert mittels Ultraschallsender und -empfänger, der Empfänger wird am Patienten befestigt. Es bietet dem Untersucher vielfältige Einsatzmöglichkeiten: Ganganalyse (Bornmann 2009), Gleichgewichtsanalyse (Vogel 2007), Funktionsanalyse des Kiefergelenks (Jüngling et al. 2004), Beweglichkeitsanalyse der HWS (Smolenski et al. 2003), der oberen Extremität (Oschmann 2010) oder des Lumbalbereichs (Will 2004) uvm. Die Messmethode mittels „Zebris“ konnte in den genannten Studien als reliabel bezeichnet werden, ist jedoch nicht für alle Untersuchungsaspekte uneingeschränkt zu empfehlen.

Zur Untersuchung der Festigkeit der Wirbelsäule gibt es ebenfalls einige Ansätze, bspw. die „Stiffness Assessing Machine“ („SAM“). Sie besteht aus einer Liege und einer Apparatur mit Messstempel, der über der Wirbelsäule angesetzt wird (Latimer 1996). Dieses Prinzip ist ähnlich dem des heutigen „Myotonometers“. Neuere Geräte nutzen elektromagnetische Unterstützung, so auch das „PAS-System“. Es wird ein elektromagnetisches Feld erzeugt und kleinste Schwankungen registriert, die einen Hinweis auf die Wirbelsäulenfestigkeit geben. Die Unterscheidung zwischen den Patientengruppen nach Geschlecht, Alter oder BMI lieferte gute Ergebnisse. Der Vergleich mit der klinischen Untersuchung war jedoch nicht zufriedenstellend (Owens 2007).

Ein weiteres modernes Verfahren ist die Messung der Schwingungsantwort eines Impulses auf das Gewebe mithilfe eines piezoelektrischen Sensors. Die elastische Verformung des Gewebes bewirkt eine elektrische Spannung, die der Sensor registriert. Daraus kann auf die Mobilität geschlossen werden (Rustler 2010). Der „Spineline“, der „ProAdjuster“, der „S.M.A.R.T.-Adjuster“ und der „PulStar“ sind mit einem piezoelektrischen Sensor ausgestattet (siehe Kapitel 1.3). Parameter und Schwingungskurve werden auf einem Monitor ausgegeben. Durch die unmittelbare Präsentation der gemessenen Werte wird eine gute Visualisierung für Therapeut und Patient geboten. Die Daten werden lückenlos aufgenommen und sind jederzeit für alle Benutzer des Geräts abrufbar. In nachfolgenden Sitzungen ist somit ein Vorher-Nachher-Vergleich durch die einheitliche Datenerhebung möglich. Das gilt ebenfalls bei einem Therapeutenwechsel, vorausgesetzt es besteht Erfahrung mit dem Instrument.

Diesen Vorteil können viele Geräte gegenüber der manuellen Untersuchung bieten. Dennoch dienen sie ausschließlich der Unterstützung der manuellen Diagnostik. Sie

dürfen nicht als Ersatz der manualmedizinischen Techniken angesehen werden, sondern als sinnvolle Ergänzung zur besseren Objektivierung einer Untersuchung. Ein Gerät sollte die Untersuchung bereichern: durch zeitökonomisches Arbeiten aufgrund schneller Messung und weniger Dokumentationsaufwand. Die Messung soll schmerzarm und wenig belastend für Arzt und Patient sein. Vor allem aber besticht ein technisches Hilfsmittel, sofern es zuverlässig misst, durch seine Objektivität. Es kann also sinnvoll sein, die traditionellen Verfahren mit einer apparativen Messmethode zu kombinieren. Um herauszufinden, inwiefern ein Instrument zur Unterstützung der Manualmedizin geeignet ist, sind im Vorfeld die Testgütekriterien zu überprüfen.

1.2 Testgütekriterien

Eine neue Testmethode für diagnostische Zwecke muss den Anspruch erheben, wissenschaftlichen Forderungen zu genügen. Dazu gehören die Hauptgütekriterien Objektivität, Reliabilität und Validität sowie die Nebenkriterien Normiertheit, Vergleichbarkeit, Ökonomie und Nützlichkeit (Lienert und Raatz 1998). Bevor Geräte wie das „TTM“ oder der „Spineliner“ in der klinischen Praxis angewendet werden, müssen zunächst die Gütekriterien überprüft werden. An dieser Stelle soll nur auf die drei Hauptgütekriterien eingegangen werden, da diese die größte Relevanz besitzen.

1.2.1 Objektivität

Die Objektivität bezeichnet den Grad, in dem die Ergebnisse einer Untersuchung/ eines Tests unabhängig von der durchführenden Person sind. Man kann drei Arten von Objektivität unterscheiden, um einen besseren Überblick über mögliche Fehlerquellen zu erhalten: Durchführungsobjektivität beschreibt die Unabhängigkeit von Verhaltensvariationen des Untersuchers während eines Tests. Auswertungsobjektivität ist die numerische oder kategoriale Auswertung nach bestimmten Regeln. Interpretationsobjektivität bedeutet, dass die Schlüsse, die aus den Auswertungsergebnissen gezogen werden, unabhängig von der interpretierenden Person sind (Lienert und Raatz 1998). Meist ist die Objektivität eines Untersuchungsverfahrens leicht zu realisieren. Es muss mithilfe einer Anleitung festgelegt werden, wie der Test durchzuführen, auszuwerten und zu interpretieren ist. Letzteres sollte mittels vorgegebenen Normen und Vergleichswerten geschehen, die

zuvor an einer repräsentativen Stichprobe von Probanden ermittelt wurden und bestenfalls nach Alter, Geschlecht o.ä. gestaffelt sind. (Bühner 2006)

Bezogen auf diese Studie, ist die Durchführungsobjektivität durch einen vorher festgelegten Messablauf gesichert. Die Auswertung übernimmt das jeweilige Gerät, indem es den empfangenen Impulsen automatisch Zahlenwerte zuordnet. Hinsichtlich der Interpretation sind Referenzwerte bezogen auf die Bevölkerung nicht relevant, da in dieser Arbeit keine Diskrimination zwischen physiologischen und pathologischen Messwerten vorgenommen, sondern ausschließlich auf die Objektivität der Messwiederholungen eingegangen werden soll.

Die Objektivität wird über die durchschnittliche Korrelation der Ergebnisse der Untersucher berechnet. Eine Korrelation von 1 ist im Optimalfall zu erreichen (Bortz 2006).

1.2.2 Reliabilität

Reliabilität wird in der Literatur auch als Zuverlässigkeit, Reproduzierbarkeit oder Präzision bezeichnet. Alle diese Begriffe beschreiben „das Maß, bis zu welchem wiederholte Messungen eines stabilen Zustandes, die durch unterschiedliche Untersucher und Instrumente, zu verschiedenen Zeiten und Stellen vorgenommen werden, zu gleichen Ergebnissen führen“ (Haerting et al. 2007).

Es gibt verschiedene Methoden, um zwei vergleichbare Messwertreihen mit einem Testverfahren zu erhalten. Der Test kann in zwei Hälften geteilt werden (Split-half-Reliabilität), es kann eine gleichwertige zweite Testversion konstruiert werden (Paralleltest-Reliabilität) oder der gesamte Test wird nach einer bestimmten Zeitspanne wiederholt (Test-Retest-Reliabilität). Der Untersucher bleibt in diesem Fall der Gleiche (Intrarater-Reliabilität), ebenso das Probandengut. Um die Messreihen mehrerer Untersucher miteinander zu vergleichen, wird die Interrater-Reliabilität bestimmt (Lienert und Raatz 1998).

Das Test-Retest-Verfahren ist für diese Studie geeignet, da mithilfe der Modelle problemlos mehrere vergleichbare Messreihen erstellt werden können, ohne dass sich die zu erwartenden Messwerte verändern. Es kann sowohl die Retest-Reliabilität innerhalb einer Sitzung (Intrasession) gemessen werden als auch zwischen zwei Sitzungen mit beliebig großem zeitlichen Abstand (Intersession).

Ein vollkommen reliabler Test ist theoretisch in der Lage, den „wahren“ Wert zu erfassen (Wirtz und Caspar 2002). Die Messtheorie besagt, dass ein „wahrer“ Wert existiert, der durch eine fehlerfreie Messung feststellbar wäre (Häder 2010). In der

Praxis tritt dieser Idealfall jedoch nicht auf, jede Messung ist mit einem, wenn auch durchaus minimierbaren, Messfehler behaftet. Die „wahren“ Werte sind unbekannt. Bei vielen Messungen geht man davon aus, dass die Messfehler um den „wahren“ Wert streuen und ihn somit eingrenzen. Man kann bei einer hohen Korrelation der Messwertreihen annehmen, dass Messwertunterschiede nicht zufällig durch Messfehler, sondern systematisch durch „wahre“ Merkmalsausprägungen bedingt sind (Bortz et al. 2006).

Die Reliabilität eines Tests gilt als ausreichend, wenn sie im Minimum 0,5 beträgt, zufriedenstellend ab einem Wert von 0,7 und hoch über 0,9 (Bortz et al. 2008). Das optimale Ergebnis ist wie bei der Objektivität ein Wert von 1.

1.2.3 Validität

Unter Validität versteht man das Ausmaß, in dem ein Test das misst, was er zu messen vorgibt. Ein Test kann nur so valide sein, wie er reliabel ist. Hohe Validität setzt somit Objektivität und Reliabilität voraus (Lienert und Raatz 1998).

Abbildung 3 zeigt mögliche Konstellationen von Reliabilität und Validität eines Messverfahrens (Haerting et al. 2007). Messungen können reliabel, aber nicht valide sein, wenn sie systematisch von dem Wert abweichen, der zu Messen beabsichtigt war (B). Messungen können im Durchschnitt valide aber nicht reliabel sein, wenn sie stark um den wahren Wert streuen (C). Im ungünstigsten Fall ist die Messung weder reliabel noch valide (D). Alle drei Konstellationen sind für ein gutes Messverfahren unbrauchbar. Ideal ist ein reliabler und valider Test (A).

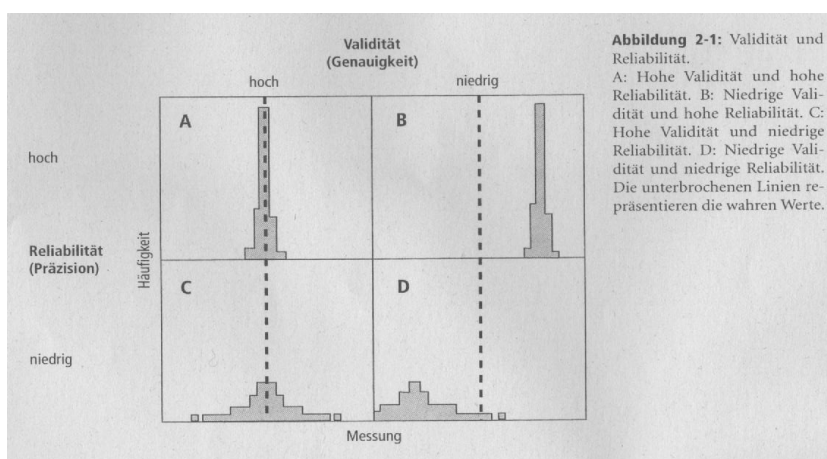


Abbildung 3: Kombinationen von Reliabilität und Validität (Harting et al. 2007)

1.3 Literaturanalyse

Wie zuvor beschrieben, muss ein neues diagnostisches Verfahren auf Objektivität, Reliabilität und Validität überprüft werden. Der folgende Abschnitt widmet sich der Suche nach bisherigen Studien über das „Tissue Tensiometer“, dem „Spineliner“ und verwandte Produkte mit anderem Markennamen. Die Durchsicht der Studien erfolgte hinsichtlich der Gütekriterien.

In einem ersten Schritt erfolgte das Zusammentragen von Schlagworten. Folgende Suchtermini bildeten die Grundlage für die Literaturrecherche: für die Thematik der Manuellen Medizin „manual medicine“, „chiropractic“, „spine dysfunction“ und „muscle tension“; für die Gütekriterien „reliability“; für Untersuchungsmethoden „test“ und „diagnosis“ und für Gerät „tool“ und „device“. Es entstanden vier Kategorien, die jeweils miteinander verknüpft wurden. Die Suche wurde in drei wissenschaftlichen Literaturdatenbanken umgesetzt. Da die Recherche einen umfassenden Einblick geben soll, wurde ein Forschungszeitraum von zehn Jahren festgelegt. Nachfolgend ist das Suchschema aufgeführt. In der Datei „Anhang1.pdf“ auf der beigefügten CD sind Tabellen mit der Anzahl der gefundenen Studien einzusehen.

manual medicine/ chiropractic/ spine dysfunction/ muscle tension
AND
reliability
AND
test/ diagnosis
AND
tool/ device

Zunächst erfolgte die Suche über „Pubmed“. „Pubmed“ ist die Datenbank der „US National Library of Medicine“. Sie beinhaltet über 21 Millionen Zitate biomedizinischer Literatur (Pubmed 2011). Die erarbeiteten Suchtermini lieferten insgesamt 614 Ergebnisse, ohne Dublikate noch 166. Für diese Arbeit waren 30 Studien relevant, welche die Objektivität oder Reliabilität eines technischen Hilfsmittels im Bereich der manualmedizinischen Diagnostik überprüften.

Die zweite Suche wurde über „Web of Knowledge“ realisiert, eine Datenbank des „Institute for Scientific Information“ (ISI). Sie vereint mehrere weltweit führende wissenschaftliche Datenbanken (Web of Knowledge 2011). Von den 179 Treffern

blieben ohne Dublikate 70 übrig, 22 Studien waren relevant. Sechs Studien wurden zuvor nicht durch „Pubmed“ gefunden.

Abschließend wurde über die „Cochrane Library“ gesucht, eine medizinische Datenbank der „Cochrane Collaboration“. Diese internationale Vereinigung von Wissenschaftlern und Ärzten stellt über 4600 Reviews bereit, die evidenzbasierte Medizin unterstützen sollen (The Cochrane Library 2011). Da nur Studien hoher methodischer Güte aufgelistet sind, wurden lediglich 21 Ergebnisse, neun ohne Duplikate, gefunden. Davon waren fünf relevant, die allesamt bereits durch die Suche über „Pubmed“ und „Web of Knowledge“ identifiziert werden konnten. Aus der hohen Überschneidung der Ergebnisse kann geschlossen werden, dass ein Großteil der relevanten Studien herausgefiltert wurde.

Zusätzlich wurde nach weiteren Schlagworten („Myotonometer“, „Spineliner“, „Pro-Adjuster“, „S.m.a.r.t.-Adjuster“) bei „Pubmed“, „Google“ oder www.springerlink.com gesucht. Die Homepages der Hersteller wurden besucht. Zudem wurden Querverweise in Studien oder dem Internet genutzt.

Unter dem Namen „Tissue Tensiometer“ sind nur zwei Geräte bekannt: eines ist im Besitz des Instituts für Physiotherapie der Uniklinik Jena und ein zweites für die Studie Buchmann et al. 2007 bereitgestellt. In der besagten Untersuchung wurde an Probanden mit triggerpunkt-assoziiertem Schmerzsyndrom die Messung der Gewebespannung mittels „TTM“ untersucht. Die palpatorisch identifizierten Triggerpunkte von zwanzig Probanden in mindestens einem von fünf Zielmuskeln wurden mit dem „TTM“ überprüft. Gegenmessungen erfolgten an Nicht-Triggerpunkten im gleichen Muskel, in zwei Vergleichsmuskeln und nach postisometrischer Relaxationsbehandlung des Triggerpunkts. Das „TTM“ konnte bei 82% der Triggerpunkte eine höhere Gewebespannung nachweisen.

Das „Tissue Tensiometer“ ähnelt dem „Myotonometer“. Zu diesem sind bereits mehrere Reliabilitätsstudien vorhanden. Leonard et al. 2003 überprüften die Reliabilität an zwei Muskeln von jeweils 35 Probanden. Dabei wurde mit unterschiedlichen Krafteinstellungen gemessen. Die Reliabilität konnte als gut bis sehr gut bezeichnet werden, vor allem bei höherer Krafteinwirkung auf den Muskel (Leonard et al. 2003). Eine weitere Studie testete das „Myotonometer“ an vier Muskeln von jeweils zehn Probanden an zwei Messtagen. Bis auf einen Muskel (schwierige Lokalisation) erreichte das Gerät bei allen eine gute bis sehr gute Test-

Retest-Reliabilität (Bizzini und Mannion 2003). Aarrestad et al. 2004 testeten das Gerät an zehn Kindern mit spastischer Lähmung. An zwei Muskeln wurde dabei jeweils in kontrahiertem und in entspanntem Zustand der gesunden und der kranken Seite von zwei Ratern gemessen. Die Reliabilität des Geräts wurde ebenfalls als gut bis sehr gut bezeichnet.

Weitere Studien wurden gesichtet - es konnte jedoch keine Forschungsarbeit gefunden werden, die das „TTM“ oder „Myotonometer“ an Modellen getestet hat. Zudem erfolgte die Überprüfung bisher ausschließlich an niedrigen Fallzahlen.

Die gleiche Funktionsweise wie der „Spineliner“ weisen die Geräte „PulStar“, „ProAdjuster“ und „S.M.A.R.T. Adjuster“ auf. Der Entwickler von „PulStar“ wirbt auf seiner Homepage mit zahlreichen selbst angefertigten Studien über Reliabilität und Therapieeffekte des Geräts, stets mit positiver Bewertung (PulStar 2011). In den wissenschaftlichen Datenbanken wurden lediglich zwei Studien über das Gerät gefunden. Darunter ist eine Anwendungsbeobachtung an zwei Individuen mit frühkindlichem Schreien aus ungeklärter Ursache. Nach nur einer Sitzung verbesserte sich die Situation bereits (Leach 2002). Die zweite Studie widmet sich der Reliabilität des „PulStars“ an 18 Probanden. Der Intrarater-ICC betrug 0,89 für den erfahrenen Untersucher und 0,78 für den neu Angelernten, die Interrater-Reliabilität lag bei 0,87 (Leach et al. 2003).

Weltweit gut vermarktet ist der „ProAdjuster“. Auf der Website des „ProAdjuster Lifestyle“ wird neben dem Gerät für eine gesunde Lebensweise und Chiropraktik im Allgemeinen geworben (ProAdjuster Lifestyle 2011). Außerdem gibt es eine „ProAdjuster“-Universität, durch die via Fernstudium der Umgang mit dem Gerät und seine Funktionsweise erlernen werden kann (ProAdjuster University 2011). Dort sind auch Hinweise zu Anwendungsbeobachtungen und einer Pilot-Studie zur Reliabilität des „ProAdjusters“ zu finden. Die Studie testet die Intra- und Interrater-Reliabilität an 64 Probanden, mit ICCs von 0,67 bis 0,86 (Tepe und Zhang, Jahr unbekannt).

Zum „S.M.A.R.T. Adjuster“ konnten keine Studien in den Literaturdatenbanken gefunden werden. Auf der Herstellerseite wird auf ein Set von vier Pilotstudien hingewiesen: zunächst der Test an einer Kalibriereinrichtung unter Laborbedingungen, anschließend die Untersuchung von Probanden mithilfe mehrerer Rater. Die Ergebnisse sind durchweg positiv (Cremonese et al. 1998). Für weitere Studien wird unter anderem auf den „Spineliner“ verwiesen (Sigma

Instruments 2011).

Das Suchwort „Spineliner“ liefert in den wissenschaftlichen Datenbanken keine Ergebnisse. Dafür konnten über die Herstellerhomepage, die Datenbank des Springerverlages und eine Broschüre zum Gerät Studien gefunden werden. Darunter befand sich eine Pilot-Studie an 51 HWS-Schmerzpatienten, mit der Frage nach Therapiewirksamkeit des „Spineliners“. Signifikante Rückgänge der Schmerzen konnten festgestellt werden (Rustler und Tilscher 2004). Eine ähnliche Anwendungsbeobachtung erfolgte an 21 Patienten mit Zervikalsyndrom, auch hier besserte sich nach Impulsgabe sowohl der palpatorische HWS-Befund als auch der goniometrische Befund des CROM. Dieser war auch 24 Stunden nach Behandlung noch nachweisbar (Zahornitzky et al. 2010). Der Hersteller wirbt mit den vielfältigsten Therapiemöglichkeiten: von segmentaler Hypomobilität und Triggerpunkten über Fersensporn und Tinnitus. Es konnten jedoch keine wissenschaftlichen Belege über die Gütekriterien oder die Wirksamkeit des Geräts gefunden werden, bis auf die Anwendungsbeobachtungen an HWS-Schmerzpatienten. Ebenso fehlt eine längerfristige Follow-up-Betrachtung, um den nachhaltigen Therapieeffekt zu beobachten.

Um erste Normwerte für die Messergebnisse des „Spineliners“ zu finden, wurden 22 männliche Probanden untersucht. Seitens des Herstellers gibt es dazu keine Hinweise, was die Arbeit mit dem „Spineliner“ erschwert. Klare Richtwerte konnten auch von Beyer et al. 2008 nicht benannt werden. Es wurde lediglich bestätigt, dass eine hohe Amplitude und eine hohe Frequenz für hartes Gewebe sprechen. Ein zweiter Versuch fand an einer Probandengruppe von 143 jungen Männern statt. Die automatisch ausgegebene Messkurve wurde genauer analysiert und in einzelne Parameter unterteilt. Für Detailaussagen seien jedoch weitere Studien nötig. Es wurde letztendlich geschlussfolgert, dass das Gerät die Objektivität einer manualmedizinischen Untersuchung anscheinend erhöht. Gleichzeitig wurde die Notwendigkeit von Reliabilitätsuntersuchungen bemerkt (Beyer et al. 2009).

Zusammenfassend gibt es nach bisheriger Recherche (Stand August 2011) keine Studien, die den „Spineliner“ an Modellen oder Probanden auf Objektivität, Reliabilität und Validität überprüft haben. Vor einem Patienteneinsatz und einer Beurteilung der zukünftigen klinischen Bedeutung des Geräts, sollten diese Untersuchungen unbedingt nachgeholt werden.

2 Ziele der Arbeit

2.1 Zielsetzung

Wie im vorhergehenden Abschnitt erörtert, sind Reliabilitätsuntersuchungen eine wichtige Voraussetzung für die Einführung eines neuen Gerätes in den klinischen Alltag. In der wissenschaftlichen Auseinandersetzung mit den Geräten „Tissue Tensiometer“ und „Spinelineer“ liegen bisher keine entsprechenden Untersuchungen vor, jedoch wurde bereits das Fehlen von Reliabilitätsstudien bemerkt (Beyer et al. 2009).

Ziel dieser Arbeit ist daher festzustellen, ob die Messergebnisse der Geräte „Tissue Tensiometer“ und „Spinelineer“ reproduzierbar und unabhängig vom Untersucher sind. Der Schwerpunkt liegt dabei auf der diagnostischen Funktion.

Nach der Skala von Bortz et al. 2008 wird eine Reliabilität als ausreichend bewertet, sofern der ICC den Wert 0,5 erreicht. Daher wurde sich bei Erstellung der Hypothesen an dem Wert von 0,5 orientiert.

Es ist nicht das Interesse der Arbeit, die Geräte miteinander zu vergleichen. Die Reliabilität an Probanden und Patienten, die Validität und mögliche Therapieerfolge zu prüfen, soll nachfolgenden Studien überlassen werden.

2.2 Fragestellungen und Hypothesen

2.2.1 Fragestellungen

- Wie hoch ist die Objektivität des Gerätes?
- Wie hoch ist die Reliabilität des Gerätes?
- Unterscheiden sich die Mittelwerte der Modelle zwischen den Modellen mit und ohne Holzbrett, zwischen den Modellen mit 8mm und 5mm dicker Deckmatte oder zwischen den Ratern?

2.2.2 Hypothesen

- Die Intrasektion-Reliabilität des Gerätes ist signifikant größer oder gleich 0,5.
H0: $ICC_{\text{Intrasektion}} < 0,5$
H1: $ICC_{\text{Intrasektion}} \geq 0,5$

- Die Test-Retest-Reliabilität des Gerätes ist signifikant größer oder gleich 0,5.
 $H_0: ICC_{\text{Retest}} < 0,5$
 $H_1: ICC_{\text{Retest}} \geq 0,5$
- Die Interrater-Reliabilität des Gerätes ist signifikant größer oder gleich 0,5.
 $H_0: ICC_{\text{Interrater}} < 0,5$
 $H_1: ICC_{\text{Interrater}} \geq 0,5$
- Die Mittelwerte der Modelle mit Holzbrett unterscheiden sich signifikant von den Mittelwerten der Modelle ohne Holzbrett.
 $H_0: \bar{x}_{\text{Modell mit Brett}} = \bar{x}_{\text{Modell ohne Brett}}$
 $H_1: \bar{x}_{\text{Modell mit Brett}} \neq \bar{x}_{\text{Modell ohne Brett}}$
- Die Mittelwerte der Modelle mit 8mm dicker Deckmatte unterscheiden sich signifikant von den Mittelwerten der Modelle mit 5mm dicker Deckmatte.
 $H_0: \bar{x}_{\text{Modell 8mm}} = \bar{x}_{\text{Modell 5mm}}$
 $H_1: \bar{x}_{\text{Modell 8mm}} \neq \bar{x}_{\text{Modell 5mm}}$
- Die durch Rater A gemessenen Mittelwerte unterscheiden sich signifikant von den durch Rater B gemessenen Mittelwerten.
 $H_0: \bar{x}_{\text{Rater A}} = \bar{x}_{\text{Rater B}}$
 $H_1: \bar{x}_{\text{Rater A}} \neq \bar{x}_{\text{Rater B}}$

Das Signifikanzniveau für die einzelnen statistischen Vergleiche wurde auf 0,05 festgelegt.

3 Material und Methoden

3.1 Studiendesign

Bei dieser Ausführung handelt es sich um eine prospektive, randomisierte Reliabilitätsstudie an Modellen.

3.2 Aufbau und Funktionsweise der Geräte

Dieser Abschnitt behandelt zunächst den Aufbau und die Funktionsweise des „Tissue Tensiometers“ und des „Spineliners“. Es folgt eine tabellarische Übersicht über diagnostische und therapeutische Nutzungsmöglichkeiten der beiden Geräte.

3.2.1 „Tissue Tensiometer“

Das in dieser Studie verwendete „Tissue Tensiometer“ (siehe Abbildung 4) ist ein Prototyp der Firma „Biofeedback Motor Control“ Leipzig.



Abbildung 4: Das „Tissue Tensiometer“

Eine Ausführung wurde eigens für die Studie durch Buchmann et al. 2007 zur apparativen Messung von Gewebespannung angefertigt. Es existieren nur die zwei Exemplare - eine serienmäßige Produktion wurde nicht geplant. Aus diesem Grund gibt es kein Handbuch für das Produkt. Nach Rücksprache mit dem Hersteller konnten einige fehlende Informationen gewonnen werden.

Das „TTM“ setzt sich zusammen aus dem Messkopf, dem Netzteil und der zugehörigen Software. Der Messkopf besteht aus einem festen „Stempel“ in der Mitte eines leicht eindrückbaren Aufsatzes (siehe Abbildung 5).



Abbildung 5: Stempel des „TTM“

Die gewonnenen Informationen werden per Infrarotschnittstelle auf den Computer übertragen. Dort können die gemessenen Werte tabellarisch dargestellt werden. Der Anwender hat somit eine gute Übersicht. Die Software ist leicht zu bedienen. Es können eigene Mess-Sets entworfen und pro Patient/ Modell gespeichert werden. Zusätzlich kann die Maximalkraft, die Messschwelle in Prozent und die Anzahl der Messwiederholungen eingestellt werden. Leider gibt es keine Anleitung zur korrekten Auswahl der Voreinstellungen.

Physikalisch gesehen kann aus Kraft pro Fläche eine Spannung berechnet werden. Das Prinzip des „TTM“ besteht darin, bei vorgegebener Eindruckkraft die zurückgelegte Wegstrecke zu messen und daraus die Gegenkraft zu berechnen. Diese Kraft, bezogen auf die Fläche des Stempels, ergibt die Spannung des Gewebes in N/m^2 .

Der Untersucher kann zunächst das Kraftmaximum bestimmen. Die Skala reicht von 1 bis 15 N. Der Computer stoppt die Messung automatisch sobald die eingegebene Maximalkraft erreicht ist. Ein Kraft-Weg-Diagramm wird auf dem Monitor ausgegeben. Der Weg ergibt sich aus der Eindringtiefe des „Stempels“ im Vergleich zum restlichen Aufsatz. Je nach voreingestellter Kraft berechnet sich eine unterschiedliche Gegenkraft pro zurückgelegtem Milimeter des Stempels.

3.2.2 „Spineliner“

Bevor mit den detaillierten Ausführungen zum „Spineliner“ begonnen wird, soll eine kurze Einführung zu den physikalischen Voraussetzungen der Messung die Grundlagen verdeutlichen.

Körper können schwingen, ebenso menschliches Gewebe. Eine einfache Schwingung ohne Reibung behält die maximale Amplitude immer bei. In der Praxis

gibt es jedoch keine reibungsfreien Schwingungen, ein Teil der Schwingungsenergie wird in eine andere Energieform umgewandelt. Deshalb nimmt die Größe der Auslenkung (maximale Amplitude) immer ab. Diesen Prozess nennt man mechanische Dämpfung (Grimsehl und Schallreuter 1991). Der „Spineliner“ misst die Dämpfungseigenschaft eines Gewebes nach Impulsgabe (Rustler 2010).

Je stärker die Dämpfung der Schwingung ausfällt, umso schneller nimmt die maximale Amplitude ab. Ein geringfügig gedämpftes System schwingt unter stetiger Amplitudenabnahme bis es den Ruhe- bzw. Ausgangszustand erreicht hat (siehe Abbildung 6, linke Kurve). Inwieweit ein Gewebe nach Impulsgabe eine solche Schwingungskurve durchläuft, ist jedoch ungewiss. Vielmehr ergibt sich die Vermutung, dass sich das Gewebe in Impulsrichtung verschiebt und sich anschließend in seinen Ursprungszustand zurückbildet, ohne in Gegenrichtung weiterzuschwingen. Diese Rückverformung könnte je nach Gewebeaufbau unterschiedlich lang andauern. Vergleichbar ist dieses Phänomen mit einem Pendel innerhalb einer Flüssigkeit: je visköser die Flüssigkeit, umso länger braucht das Pendel, um nach Impulsgabe zu seinem Ausgangspunkt zurückzuschwingen. Wenn die Dämpfung zunimmt, so gibt es einen Punkt, an dem zum ersten Mal eine solche aperiodische Schwingung (siehe Abbildung 6, rechte Kurve) auftritt. Diesen Fall nennt man Kriechfall.



Abbildung 6: Schwingungskurve bei schwacher (li) und starker (re) Dämpfung (Grimsehl und Schallreuter 1991)

Die aperiodische Schwingungskurve ist mit der vom „Spineliner“ ausgegebenen Kurve vergleichbar. Da keine exakten physikalischen Größenangaben in den Diagrammen angegeben sind, kann die Existenz einer aperiodischen Schwingung oder eines Kriechfalls nur an Hand der qualitativen Form der Kurve beurteilt werden. Wichtig ist dafür insbesondere eine aperiodische Näherung an die t -Achse und damit

verbunden ein Wechsel der Krümmung vom Maximum an.

Neben der Kurve werden die Messwerte Widerstand, Mobilität, Frequenz und Wellenform ausgegeben. Der Widerstand entspricht der maximalen Amplitude (peak), die Frequenz dem Kehrwert der Schwingungsdauer ($2 \cdot \text{time total}$), die Mobilität der Mobility ($\text{time to peak} / \text{time total} \cdot 100$) und die Wellenform der Motoricity ($\text{area to peak} / \text{area total} \cdot 100$). Beyer et al. 2009 haben diese und weitere Parameter anhand der Kurve grafisch dargestellt (siehe Abbildung 7).

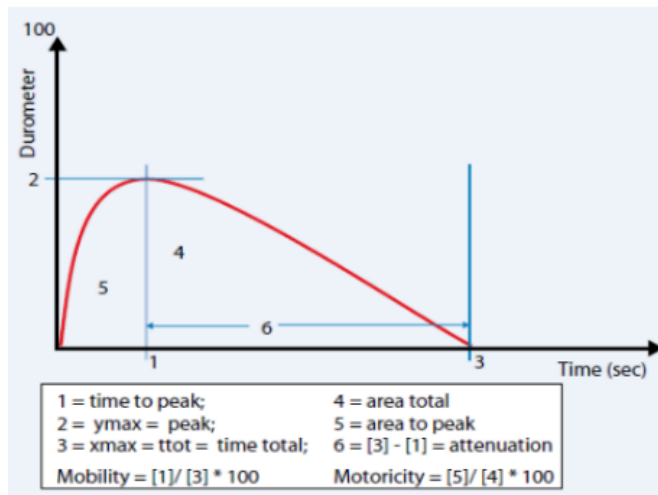


Abbildung 7: Parameter der Schwingungskurve (Beyer et al. 2009)

Feste Strukturen schwingen mit höherer Frequenz und Amplitude als weiche Strukturen. Übertragen auf die Manualmedizin bedeutet eine hohe Amplitude und Frequenz eine Hypomobilität eines Gelenks bzw. Tonuserhöhung eines Muskels, analog zeigen eine niedrige Amplitude und Frequenz eine Hypermobilität bzw. Tonusminderung. Diese soll der „Spineliner“ erkennen.

Das Gerät sendet einen Impuls an das zu untersuchende Gewebe und nutzt einen piezoelektrischen Sensor, um die entstehenden Schwingungen aufzuzeichnen. Dieses System wird bereits seit Jahren in Raumfahrt-, Luftfahrt- und Brückenbautechnik zur Bestimmung von Verarbeitungsfehlern, wie Rissen im Material, verwendet. Die Idee, dieselbe Technik bei der Untersuchung von Patienten zu gebrauchen, ist der Grundgedanke des „Spineliners“. Erste Entwicklungen hinsichtlich dieses Vorhabens begannen bereits im Jahr 1987, zehn Jahre später ging es in Produktion (Rustler 2010).

Im Messkörper befinden sich neben dem piezoelektrischen Sensor ein Drucksensor, Spule, Hammer und Amboss (siehe Abbildung 8).

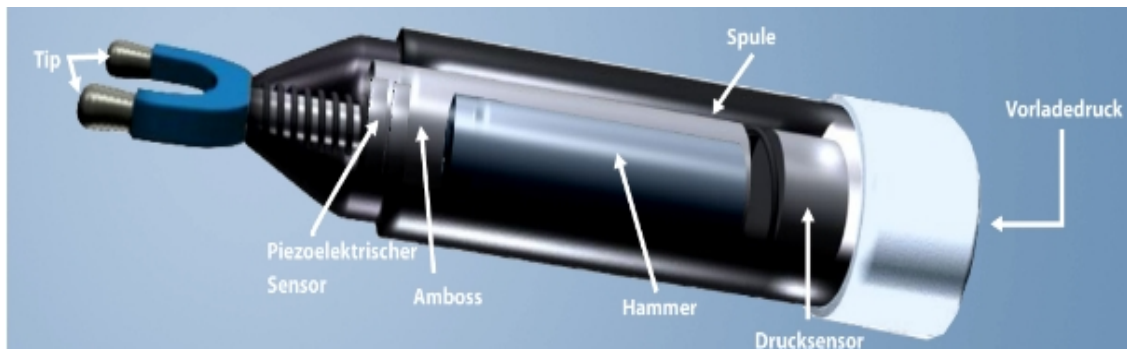


Abbildung 8: Aufbau des „Spineliners“ (TTC 2008)

Der Untersucher umgreift das Gerät und setzt es an der zu messenden Oberfläche auf (Abbildung 9).

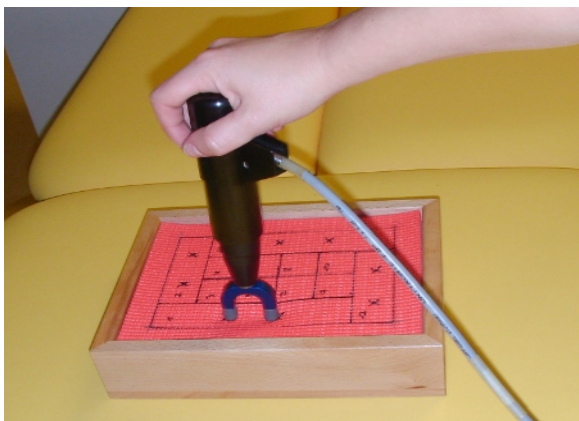


Abbildung 9: Handhabung des „Spineliners“

Mithilfe des Handballens wird eine Vorspannung von ca. sechs Pfund (2,7 kg) überwunden. Dadurch ist der notwendige Tiefenkontakt von Gerät zu Gewebe sichergestellt. Der Drucksensor registriert die Vorlast und generiert daraufhin einen elektrischen Impuls, der durch die Spule in kinetische Energie umgewandelt wird. Somit ist die Kraft, die auf die Körperstrukturen übertragen wird, objektiv, d.h. unabhängig vom Benutzer immer die gleiche. Über Hammer, Amboss und Sensorkopf wird der kinetische Impuls schließlich auf das Messobjekt übertragen (Rustler 2010).

Es werden vom Hersteller verschiedene Aufsätze zur Variation der Mess- bzw. Therapieoberfläche angeboten (Abbildung 10). In dieser Studie wurde der Standard-Applikator für die Wirbelsäule verwendet, mit einem Abstand der beiden Tips von 30mm.



Abbildung 10: Auswahl der Behandlungsaufsätze (Rustler 2010)

Im klinischen Alltag wird der Sensorkopf im 45°-Winkel (nach kaudal) auf Facettenebene jedes einzelnen Wirbelsegments aufgesetzt, durch einen speziell angepassten Untersuchungsstuhl wird die Position des Patienten optimiert.

Der elektromechanisch generierte Impuls mit einer Stärke von 421,84 g/cm sorgt für eine Schwingung des angrenzenden Gewebes. Mit dem piezoelektrischen Sensor wird dessen Dämpfung registriert und an den Computer weitergeleitet.

Die mitgelieferte Software verarbeitet die Signale und bereitet sie benutzerfreundlich auf. Es wird die Schwingungskurve und vier Parameter (Widerstand, Mobilität, Frequenz, Wellenform) angezeigt. Das ermöglicht eine übersichtliche Dokumentation von Diagnostik und Therapie. Für jeden Patienten kann eine eigene Akte mit mehreren Fällen und jeweils chronologischer Auflistung der Untersuchungszeitpunkte angelegt werden. Wird ein Fall aufgerufen, können die gemessenen Parameter tabellarisch oder grafisch dargestellt werden, oder direkt an einen Drucker gesendet werden.

3.2.3 Übersicht

Tabelle 1: Übersicht der Eigenschaften des „TTM“ und des „Spineliner“

	„Tissue Tensiometer“	„Spineliner“
Aufbau	Messkopf, Software	Messkopf, Software
Wirkungsweise	„Stempel“ misst Eindrucktiefe, Berechnung der Gewebespannung	piezoelektrischer Sensor misst Dämpfung der Schwingung
Diagnose	Identifikation von erhöhter Gewebespannung (z.B. Triggerpunkte)	Identifikation von segmentalen Funktionsstörungen oder erhöhter Gewebespannung
Therapie	keine	durch wiederholte Impulsgabe möglich
Anwendung	bisher nur in Studien	weltweit, v.a. bei Patienten mit chronischen Muskel- und Gelenkbeschwerden
bisherige Studien	- Buchmann et al. 2007/ Objektivierbare apparative Messung der Gewebespannung mittels TTM	- Rustler und Tilscher 2004/ Treatment of the cervical spine with the Spineliner - Beyer et al. 2008/ Eigenschaften spinaler segmentaler Beweglichkeit - Beyer et al. 2009/ Beweglichkeit der Wirbelsäulensegmente - Zahornitzky et al. 2010/ Behandlung des oberen Zervikalsyndroms mit funktioneller computer-assistierter Reflextherapie (Spineliner)

3.3 Gewebemodelle

Zunächst soll kurz erläutert werden, warum diese Studie an Modellen statt an Probanden durchgeführt und wie diese Idee umgesetzt wurde. Der menschliche Körper unterliegt ständigen Veränderungen: durch Tagesform, Stimmung, Trainingszustand, Schmerzen, Krankheiten und vieles mehr. Das Gewebe reagiert empfindlich auf diese Einflüsse. Die Gewebespannung kann bei Stress erhöht und ein paar Tage später (bspw. nach einem langen Wochenende) wieder normalisiert sein. All das ist sehr schwer zu eruieren und kann die Ergebnisse einer Reliabilitätsmessung am Patienten beeinflussen.

Aus diesem Grund haben sich in der Praxis Modelle als ideal erwiesen, um annähernd nur die Geräteeigenschaften zu testen. Modelle sind nahezu unveränderlich und stehen unbegrenzt für Messungen zur Verfügung. Die Überlegung, Modelle mit einem breiten Spektrum an Härtegraden herzustellen, wurde zugunsten einer realitätsnäheren Variante verworfen. Es ist anzunehmen und zu erwarten, dass die Geräte zwischen sehr weichen Strukturen (bspw. einem Kissen) und besonders harten Gegenständen (bspw. einem Stein) gut differenzieren können. Wichtiger ist die Frage, ob das Gerät Strukturen unterscheiden kann, die sich grundsätzlich sehr ähnlich sind und nur feine Unterschiede aufweisen. Die Gewebeschichten der verschiedenen Menschen sind im Grunde immer gleich aufgebaut, jedoch durch Einflüsse wie Sonnenstrahlung, Ernährung, Sport etc. geprägt. Es galt also, Modelle mit einem gleichen strukturellen Aufbau aber Unterschieden in der jeweiligen Ausprägung der Schichten zu entwickeln.

Für die Umsetzung der Modelle wurden zwanzig Holzkästen in zehn verschiedenen Größen mit je zweifacher Ausführung hergestellt. In diese konnten anschließend verschiedene Matten und Schaumstoffe als Gewebeersatz eingelegt werden (siehe Abbildung 11).



Abbildung 11: Aufbau eines Gewebemodells

Ziel war es, sich möglichst nah an der Realität des menschlichen Gewebes am Rücken (Wirbelsäule) zu orientieren, das auch in Natura unterschiedliche Ausmaße aufweist. Auf jedem dieser Modelle befinden sich zwölf Messpunkte, die die verschiedenen Bewegungssegmente repräsentieren sollen. Dies entspricht der Anzahl der Wirbelkörper der BWS. Abbildung 12 zeigt am Beispiel des Modell 3A die Markierungen auf der Deckmatte. Das „TTM“ wurde jeweils bündig zum linken oberen Rand jedes Kästchens aufgesetzt. Der „Spineliner“ wurde entweder mit den Tips auf den Kreuzen (wenn vorhanden) oder in den quadratischen Kästchen mit einem Tip oben links und dem anderen Tip unten rechts aufgesetzt.

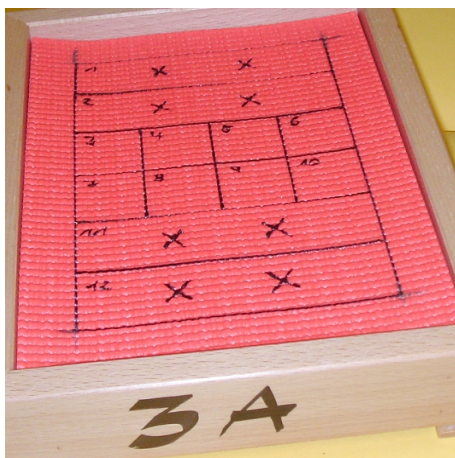


Abbildung 12: Deckmatte des Modell 3A mit Markierungen

Mittels Losentscheid wurde für jedes Modell ermittelt, welche Maße der Holzkasten erhält, ob die Hautmatte 5 mm oder 8 mm messen soll, ob die Fettschicht aus weichem oder härterem Schaumstoff besteht, ob die Muskelmatte 1 cm oder 1,5 cm dick wird und ob ein Holzbrett stellvertretend für eine Verhärtung eingelegt wird.

Als Hautmatten wurden die Gymnastikmatte „Yoga“ (5 mm dick) von „Sportco“ und die Gymnastikmatte „Komfort“ (8 mm dick) von „Eyblsport“ verwendet. Für die Fettschicht wurde Schaumstoff „1a (weich)“ und „1b (mittel)“ von einem Raumausstattungsunternehmen zur Verfügung gestellt. Um die Muskelschicht zu repräsentieren, dienten die „Energetics Gymnastikmatte Sport Body Fit XL“ von „Sportolino“ (1 cm dick) und die „Yogamatte“ von „D&S Vertriebs GmbH“ (1,5 cm dick). Das für die Hälfte der Modelle als „Verhärtung“ genutzte Holzbrett ist 1 cm dick. Die Benennung der Modelle erfolgte durch Nummerierung von 1 bis 10 (gleiche Nummer bedeutet gleiche Maße des Holzkasten) mit jeweiliger Unterteilung in A und

B – entweder mit oder ohne Funktionsstörung in Form einer Holzschicht. Dabei ist die Zuordnung von A/B und Verhärtung/nicht Verhärtung bewusst zufällig gewählt, damit der Untersucher nicht vom Modellnamen bereits ableiten kann, ob sich ein Holzbrett darin befindet oder nicht. In Tabelle 2 ist für jedes Modell die Mattendicke und Schaumstoffhärte aufgelistet und ob ein Holzbrett (Verhärtung) zugeteilt wurde. Ein vorhandenes Brett ist durch „+“, Modelle ohne Brett durch „-“ gekennzeichnet.

Tabelle 2: Modelle mit Zuordnung der Schichten

Modell	„Hautdicke“ in mm	„Fett“	„Muskeldicke“ in cm	„Verhärtung“
1A	8	weich	1,0	-
1B	8	weich	1,5	+
2A	8	mittel	1,5	-
2B	5	mittel	1,5	+
3A	5	mittel	1,0	+
3B	5	mittel	1,5	-
4A	5	weich	1,5	-
4B	5	weich	1,0	+
5A	5	mittel	1,5	+
5B	8	weich	1,0	-
6A	8	mittel	1,0	+
6B	5	mittel	1,5	-
7A	5	mittel	1,0	-
7B	8	weich	1,5	+
8A	5	mittel	1,0	-
8B	5	mittel	1,0	+
9A	5	mittel	1,5	+
9B	8	mittel	1,0	-
10A	5	weich	1,5	+
10B	8	weich	1,5	-

Auf jeder „Hautschicht“ der insgesamt 20 verschiedenen Gewebeimitationen wurden die zwölf Messpunkte markiert und durchnummeriert, so dass bei jeder Untersuchung der exakt selbe Punkt in der gleichen Reihenfolge gemessen wird, also die gleichen Bedingungen herrschen. Eine weniger exakte Definition der Messpunkte könnte die Fehlerrate erheblich steigern oder sogar verdoppeln (Haneline 2009).

Zwei Modelle mit dargestellten Schichten sind in Abbildung 13 auf der nächsten Seite zu sehen.

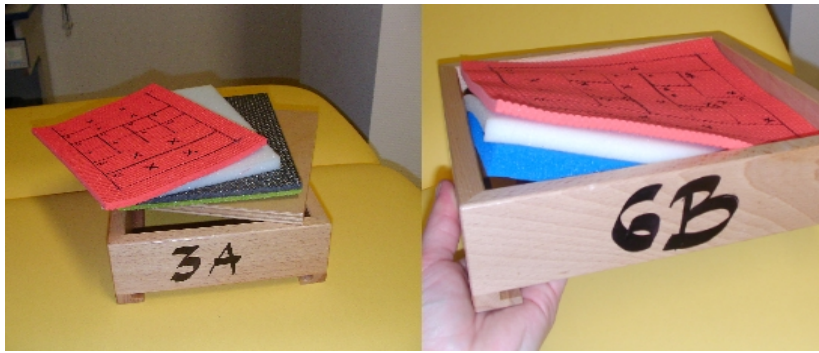


Abbildung 13: Zwei Modelle, Schichten dargestellt

Es erfolgte für weitere Fragestellungen, neben der Berechnung der Objektivität und Reliabilität, eine Einteilung der Modelle nach ihrer Deckmatte (8 mm oder 5 mm dick) und nach dem Vorhandensein eines Holzbretts. Die obere Matte wurde durch den Stempel der Geräte sichtbar eingedrückt, während das Holzbrett eine durch alle Schichten reichende Verhärtung darstellt. Durch das Prüfen dieser beiden Kriterien kann abgeschätzt werden, ob das Gerät durch die obere Gewebeschicht (die Haut) bereits beeinflusst wird, oder ob es auch tiefliegendere Veränderungen unabhängig von der Deckschicht erfassen kann.

Eine Unterteilung mit Berücksichtigung aller Modellschichten war aufgrund der hohen Vielfalt nicht sinnvoll. Außerdem wurden durch die zufällige Zuweisung der Merkmale nicht alle möglichen Merkmalskombinationen abgedeckt, bzw. wären pro Kombination nicht genügend Modelle vorhanden. In nur fünf Fällen haben zwei Modelle den gleichen Aufbau, in einem Fall drei Modelle.

3.4 Studienablauf

Zu Beginn wurden die Rater mit der bis dato unbekannten Methode in einer kurzen Trainingsphase vertraut gemacht. Fehlerquellen konnten beseitigt werden und es wurde sichergestellt, dass die Handhabung des Geräts zwischen Rater A und Rater B übereinstimmt. Die Vorgehensweise bei den jeweiligen Messungen wurde gerätespezifisch vereinbart, um für höchstmögliche Standardisierung zu sorgen. Die Intrasession-Messung wurde von einem Rater an zehn zufällig gewählten Messpunkten durchgeführt.

Für jedes Gerät wurden fünf Messtermine vereinbart. Dazwischen lag je ein Abstand von einer Woche, um eine übermäßige Gewöhnung eines Untersuchers an das

Gerät zu verhindern (siehe Abbildung 14).

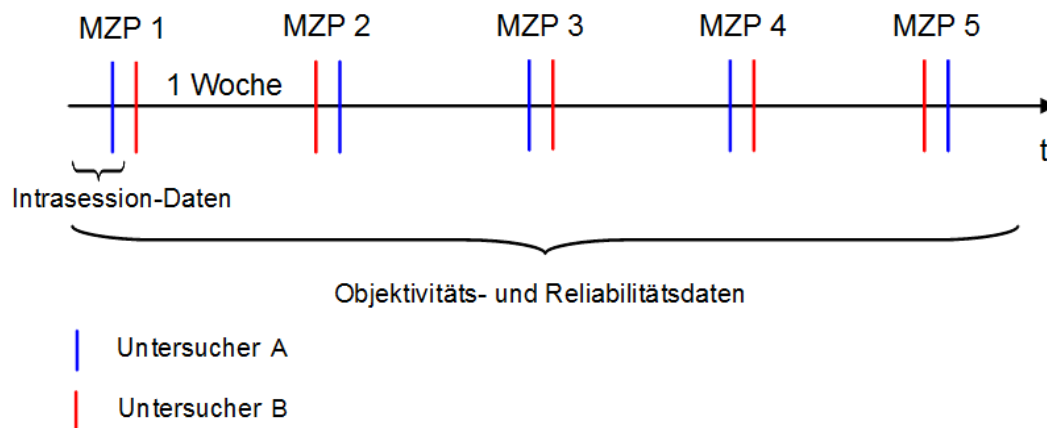


Abbildung 14: Studienablauf

Vor Beginn eines jeden Messzyklus wurden an drei Modellen Probemessungen durchgeführt, um den Messablauf zu rekapitulieren. Dazu kontrollierte der jeweils nicht messende Rater den gerade messenden Rater. Durch die gegenseitige Kontrolle konnte die standardisierte Durchführung der Messung konstant gehalten werden. Für jeden Messzeitpunkt wurde eigens ein Messprotokoll angefertigt. Die darauf festgeschriebene Reihenfolge der Modelle und MZIP der Rater pro Modell wurde zufällig computergestützt generiert (www.randomization.com). Damit sollten etwaige systemische Fehler durch eine eventuelle anfängliche Unsicherheit und spätere Ermüdungserscheinungen auf alle Modelle gleichermaßen aufgeteilt werden.

3.4.1 Messung mit dem „Tissue Tensiometer“

Die Messungen fanden in einem Untersuchungsraum des Instituts für Physiotherapie in der Uniklinik Jena statt. Eine Untersuchungsfläche diente als Unterlage für die Modelle, dabei wurde auf die stets gleiche Position in Bezug zur Fläche geachtet (Modell genau mittig). Der Laptop mit der benötigten Software stand neben dem Modell (siehe Abbildung 15).

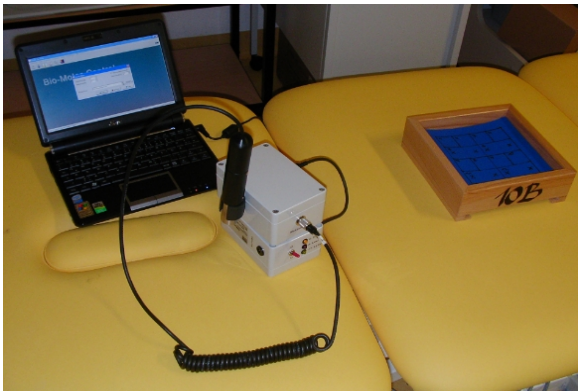


Abbildung 15: Untersuchungsaufbau „TTM“

Das Programm bietet verschiedene Einstellungen, um eine Messung zu definieren. Im Vorfeld wurden eine einmalige Messung pro Messpunkt, eine Kraft von sechs N und eine Messschwelle von zehn Prozent als Messparameter festgesetzt.

Vor Liege und Modell befand sich der messende Rater und setzte den Messstift genau senkrecht in die rechte untere Ecke der aufgezeichneten Kästchen, so dass die Linien der Kästchen mit dem äußeren Umfang des Geräts abschlossen. Der Druck wurde vom Messenden langsam gesteigert. An der auf dem Bildschirm dargestellten Kurve konnte die Eindringtiefe pro angewendeter Kraft verfolgt werden. Sobald der voreingestellte Maximaldruck von sechs N, also das Ende der x-Achse, erreicht war, wurde über ein Dialogfenster die Richtigkeit der Messung abgefragt, mit dem „OK-Button“ bestätigt und damit die Messung beendet (siehe Abbildung 16).

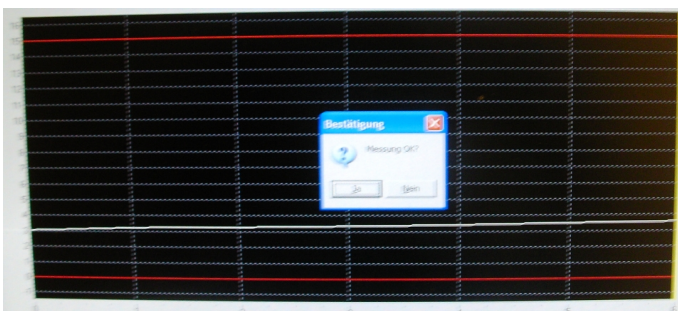


Abbildung 16: Bildschirmausgabe nach einer Messung mit dem „TTM“

Eine zweite Person war somit nötig, um das Programm während der Messung zu bedienen. Die Messung wurde als „OK“ gewertet, wenn eine leicht von links nach rechts ansteigende Kurve (weiße Linie in Abbildung 16) zu sehen war, also Kraft und Eindringtiefe gleichmäßig zunahmen. Inakzeptabel waren „Zick-Zack-Kurven“,

Kurven im Minusbereich, Kurven die am linken Rand endeten, Kurven die nicht am linken Rand begannen oder leere Diagramme. Dann musste der Messpunkt wiederholt werden.

Da pro Serie beliebig viele Punkte gemessen werden können, musste ein eigenes Messpunktset erstellt werden. Es wurde sich für ein 24-Punkt-Set entschieden, um in einem Zug ein Modell von beiden Ratern durchmessen zu lassen. Danach konnten die Werte gespeichert oder gedruckt werden. Mithilfe des Messprotokolls war die Reihenfolge der Rater im Nachhinein reproduzierbar. Das Ergebnis für die Gewebespannung eines jeden Messpunktes wurde in eine Datenbank übertragen.

3.4.2 Messung mit dem „Spineliner“

Wie bereits im vorhergehenden Abschnitt beschrieben, wurden die Modelle auf einer Untersuchungsliege mittig positioniert. Daneben befand sich der Computer für den „Spineliner“. Der Messende stand mit dem Rücken zum Monitor, um Beeinflussungen durch das sichtbare Messergebnis zu vermeiden (siehe Abbildung 17).



Abbildung 17: Untersuchungsaufbau „Spineliner“

Es wurde der Standard-Applikator für die Wirbelsäule verwendet, mit einem Abstand der beiden Tips von 30mm. Dieser Abstand entspricht auch dem diagonalen Durchmesser der Quadrate, die auf den Modellen eingezeichnet sind. Die Messpunkte wurden immer mit der gleichen Ausrichtung der Tips gemessen: ein Tip links oben im Quadrat, der andere rechts unten. Dabei wurde sich geeinigt, das Anschlusskabel immer links zu halten und das Gerät nach der Messung direkt an den anderen Rater zu übergeben, da sich die Tips beim Ablegen des Geräts verdrehen können.

Das Aufsetzen der Tips erfolgte nicht im für die Wirbelsäule empfohlenen 45°-Winkel, sondern im 90°-Winkel, da dieser leichter visuell zu überprüfen und beim Abdrücken des Geräts einzuhalten ist. Es wurde nur mit leichtem Druck gearbeitet, um gerade so die Vorspannung, die mithilfe der Software auf „1“ eingestellt war, zu überwinden. Wurde ein Punkt gemessen, ging das Gerät ohne Abfrage zum nächsten Punkt über. Ein akustisches Signal bestätigte den Erfolg einer Messung sowie das Ende der Messreihe, nachdem die Tonhöhe kontinuierlich maximal angestiegen war. Es wurde also kein zweiter Rater zum Bedienen des Programms während einer Messung benötigt.

Parallel zur Messung des einen Raters schrieb der jeweils andere Rater die aktuelle Uhrzeit auf das Messprotokoll. Diese Maßnahme sollte vor allem zur Kontrolle dienen, ob ein Modell bereits gemessen wurde oder falls es ein Problem in einer Messreihe gab.

Da die Rater-Bezeichnung leider nicht mit den Daten zusammen ausgedruckt werden konnte, musste mithilfe des Protokolls die Untersucherreihenfolge noch einmal nachvollzogen und markiert werden (Rater A pink, Rater B grün, siehe Abbildung 18).

Erstbesuch			15/04/2010 04:50:23			Letztbesuch			15/04/2010 04:52:32		
			Behandlungen			Behandlungen					
C3	Erste	Letzte									
Widerstand											
Mobilität											
Frequenz											
Wellenform											
C7	Erste	Letzte									
Widerstand											
Mobilität											
Frequenz											
Wellenform											
T4	Erste	Letzte	T5	Erste	Letzte	T6	Erste	Letzte			
Widerstand	40	46	Widerstand	37	35	Widerstand	39	37			
Mobilität	57	56	Mobilität	56	55	Mobilität	58	60			
Frequenz	37	39	Frequenz	37	39	Frequenz	38	39			
Wellenform	54	51	Wellenform	54	51	Wellenform	56	56			

Abbildung 18: Ausdruck einer „Spineliner“-Messung

Der Ausdruck erfolgte modellweise. Für jedes Modell wurde eine eigene virtuelle Akte angelegt, so wie auch später mit einem Patienten zu verfahren ist. Ruft man das jeweilige Modell im Programm auf, so kann man den Messzeitpunkt (=„Fall“) auswählen und sich von diesem die verschiedenen „Besuche“, hier je einmal von Rater A und einmal von Rater B, anzeigen lassen. Die Software stellt auch die Uhrzeit des Besuchs dar, die noch einmal vor dem Drucken mit der Uhrzeit auf dem Protokoll verglichen wurde.

Wie ebenfalls in Abbildung 18 zu sehen ist, gibt die Software die Messpunkte mit der Bezeichnung „C1A-C7“ (HWS), „T1-T12“ (BWS) und „L1-S1“ (LWS) vor. Für diese Untersuchung war die Bezeichnung der Wirbelkörper allerdings irrelevant. Die drei Gruppen waren separat messbar. Allerdings mussten je Gruppe immer alle Punkte gemessen werden, damit das Programm die Daten speichert. Da diese Studie pro Modell zwölf Messpunkte eingeplant hatte, war es gut möglich, die Gruppe „T1-T12“ zu nutzen. Von den ausgegebenen Werten wurden alle vier Parameter genutzt und zur Interpretation herangezogen, um eine aussagekräftige Datengrundlage zu erhalten.

3.5 Statistische Methoden

Bei der Berechnung der ICCs über „SPSS Statistics“ konnte die Zuordnung der jeweils zwölf Messpunkte zu den zwanzig Modellen nicht berücksichtigt werden. Daher wurde jeder Messpunkt als ein eigenes Modell gewertet. Somit ergaben sich 240 „Punktmodelle“.

Reliabilitätsmessungen werden üblicherweise mithilfe des Intraclass Correlation Coefficient (ICC) beurteilt. Dabei wird nicht nur die Richtung der Variablen verglichen, sondern auch ihr mittleres Niveau (Bühl 2008). Die Einschätzung der Reliabilität anhand der ICC-Werte erfolgt nach Bortz et al. 2008 (siehe Tabelle 3). In Anlehnung daran werden auch die Werte der Objektivität interpretiert.

Tabelle 3: Interpretation der ICC-Werte im Hinblick auf die Reliabilität (Bortz et al. 2008)

ICC	Wertung
> 0,9	hoch
0,7 – 0,9	zufrieden stellend
0,5 – 0,7	ausreichend
< 0,5	nicht ausreichend

Die Auswertung der Ergebnisse erfolgte unter Zuhilfenahme von „SPSS Statistics“ und des Tabellenkalkulationsprogramms „Excel“. Grafiken wurden mithilfe von „SPSS Statistics“ und des Schreibprogramms „Word“ erstellt. Das „Tissue Tensiometer“ lieferte einen Parameter, die Gewebespannung. Der „Spinelineer“ gab vier Parameter einer Messung aus: den Widerstand, die Mobilität, die Frequenz und die Wellenform. Für jede dieser Variablen wurde der ICC der Intrasession-, Test-Retest- und

Interrater-Reliabilität berechnet.

Die Objektivität wurde als Korrelation nach Bravais und Pearson zwischen Rater A und Rater B für jeden Messzeitpunkt berechnet. Es handelt sich um ein Maß für den Grad des linearen Zusammenhangs zweier Merkmale. Bei einem Wert von 1 bzw. -1 besteht der größtmögliche positive bzw. negative Zusammenhang zwischen den Variablen, bei einem Wert von 0 besteht kein Zusammenhang (Hartung et al. 2005).

Die Berechnung des ICCs erfolgte für die Intrasession-Reliabilität mit dem Modell einfach-zufällig, da der Rater (in diesem Fall Rater A) festgelegt war und nur die Modellreihenfolge zufällig gewählt wurde. Für den Vergleich der Test-Retest- und Interrater-Reliabilität zweier Messzeitpunkte wurde das Modell zweifach-zufällig (Rater ebenfalls zufällig) verwendet. Gewertet wurden „Korrelation innerhalb der Klasse“, „Einzelne Maße“, da nicht die Mittelwerte, sondern alle Werte einzeln miteinander verglichen werden sollen. Orientiert wurde sich dafür an Wirtz und Caspar 2002, Shrout und Fleiss 1979. Das Statistikprogramm der Firma „SPSS“ wurde auch in anderen Reliabilitätsstudien verwendet, bspw. bei Lidström 2009 und Kachingwe 2005.

Der Abgleich der ermittelten ICCs mit dem Grenzwert der Hypothesen von 0,5 konnte im Zuge der ICC-Berechnung mithilfe „SPSS Statistics“ erfolgen. Das Programm gibt einen p-Wert aus, nachdem der Prüfwert (in diesem Fall 0,5) und das Signifikanzniveau (Standard 0,05) definiert wurde. Unterschreitet der p-Wert die 0,05, so liegt ein statistisch signifikantes Ergebnis vor. Dies bedeutet, dass die Wahrscheinlichkeit, eine eigentlich richtige Nullhypothese irrtümlich abzulehnen, weniger als 5% beträgt.

Zur Berechnung der Overall-Werte (über alle 5 Messzeitpunkte) der Test-Retest- und der Interrater-Reliabilität waren komplexere Formeln notwendig, die in Eliasziw et al. 1994 beschrieben wurden. Dies ist der Tatsache geschuldet, dass die Berechnung durch „SPSS Statistics“ entweder für den Vergleich eines MZP zwischen zwei Ratern geeignet ist, oder aber für den Vergleich von mehreren MZP von nur einem Rater. Es ist zwar möglich, alle fünf MZP beider Rater in die Berechnung einfließen zu lassen, allerdings werden dann die MZP eines selben Raters nicht als solche erkannt, sondern als MZP fünf verschiedener Rater.

Die von Eliasziw et al.1994 beschriebenen Formeln berücksichtigen die Anzahl der MZP pro Rater. Abbildung 19 zeigt beispielhaft die Formel der Berechnung des Interrater-ICC unter der Bedingung, dass die Rater zufällig ausgewählt werden können:

$$\hat{\rho}_{inter, random} = \frac{cov(x_{ijk}, x_{ij'k})}{var(x_{ijk})} = \frac{\hat{\sigma}_S^2}{\hat{\sigma}_S^2 + \hat{\sigma}_R^2 + \hat{\sigma}_{SR}^2 + \hat{\sigma}_e^2}$$

Abbildung 19: Formel zur Berechnung der Interrater-ICCs für zufällige Rater (Eliasziw et al.1994)

Um die, für die Formel benötigten, Varianzkomponenten zu berechnen, wurden zunächst die mittleren Quadrate für das Subject (Modell), den Rater, die Subject-Rater-Interaktion und den erwarteten Fehler per „SPSS Statistics“ ermittelt. Diese Werte konnten dann in weitere Formeln eingesetzt werden, die bei „Excel“ eingegeben wurden.

Zur Beurteilung, ob die manuell berechneten ICC-Werte signifikant über 0,5 liegen, wurden durch Eliasziw et al.1994 ebenfalls eigene Formeln angegeben. Das jeweilige Ergebnis (F_{inter} oder F_{intra}) konnte dann mit einem zuvor ermittelten Grenzwert verglichen werden. Dieser Grenzwert berücksichtigt die Anzahl der Modelle (hier 240) und beträgt für die Interrater-Reliabilität 1,237 und für die Intrarater-Reliabilität 1,178. Liegen die F-Werte über diesem Grenzwert, kann von einem signifikant (das Signifikanzniveau beträgt auch hier 0,05) über 0,5 liegenden ICC-Wert ausgegangen werden.

Für den Vergleich zwischen den einzelnen Gruppen (Deckmatte 8 mm oder 5 mm dick, Holzbrett ja oder nein, Rater A oder B) war ein einfacher t-Test ungeeignet, da alle Faktoren am gleichen Datenpool überprüft werden sollten. Somit wurde eine Varianzanalyse mit Messwiederholungen nach dem „ANOVA“-Modell vorgenommen, wodurch eine vollständige Ausnutzung der gemessenen Information möglich ist. Diese geht von normalverteilten Daten aus, daher wurde zuvor auf Normalverteilung getestet.

Um die exakte Auswahl der statistischen Methoden und deren korrekte Ausführung sicherzustellen, erfolgte eine Beratung durch einen Mitarbeiter des Instituts für

Medizinische Statistik, Informatik und Dokumentation der Friedrich-Schiller-Universität Jena sowie die Vorstellung der Studie in einer Teambesprechung.

4 Ergebnisse

Im folgenden Kapitel werden die Ergebnisse der Datenverarbeitung präsentiert. Es werden Mittelwertvergleiche angestellt, um bspw. Aussagen über die Eignung der Modelle treffen zu können (nachzulesen in Kapitel 5). Objektivität und Reliabilität der Geräte werden in Bezug zu den Hypothesen gesetzt. Ebenso werden die Mittelwertunterschiede zwischen den zwei verschiedenen Deckmatten, den Modellen mit oder ohne Holzbrett und den beiden Ratern geprüft, um Aussagen hinsichtlich der Differenzierungseigenschaften der Geräte zu formulieren.

Die Berechnungen erfolgten wie in Kapitel 3.5 beschrieben.

4.1 „Tissue Tensiometer“

4.1.1 Mittelwertvergleiche

Beim Vergleich der Mittelwerte der einzelnen Modelle und Rater über alle Messzeitpunkte (MZP) ist zu erkennen, dass Rater A insgesamt niedrigere Werte gemessen hat als Rater B, außer bei Modell 9a und Modell 10b. Eine ausführliche Mittelwerttabelle mit den berechneten Standardabweichungen ist nachfolgend aufgeführt (Tabelle 4).

Tabelle 4: Mittelwerte des „TTM“ pro Rater über alle MZP in N/m² (n=60 pro Modell, n=1200 über alle Modelle)

Modell	Mittelwert Rater A	SD Rater A	Mittelwert Rater B	SD Rater B
1a	3,28	0,68	3,48	0,92
1b	3,52	1,66	3,67	1,43
2a	3,07	0,99	3,29	1,39
2b	2,40	0,87	2,46	0,84
3a	2,96	0,85	3,06	0,95
3b	2,47	0,62	3,01	1,75
4a	2,54	0,75	2,62	0,59
4b	2,77	0,70	3,37	1,25
5a	2,70	1,11	2,94	1,95
5b	3,51	1,76	3,84	2,00
6a	3,51	0,67	3,54	1,07
6b	2,57	0,91	2,64	0,89
7a	2,93	0,90	3,01	1,23
7b	3,33	1,11	3,59	2,02
8a	2,47	0,60	2,68	0,86

8b	3,02	1,69	3,35	1,67
9a	2,95	1,09	2,40	0,47
9b	3,13	0,98	3,39	1,40
10a	3,03	1,05	3,42	1,86
10b	3,59	1,61	3,40	1,39
Alle Modelle	2,99	1,03	3,16	1,30

Zur Visualisierung der Messwertunterschiede wurde Diagramm 1 erstellt.

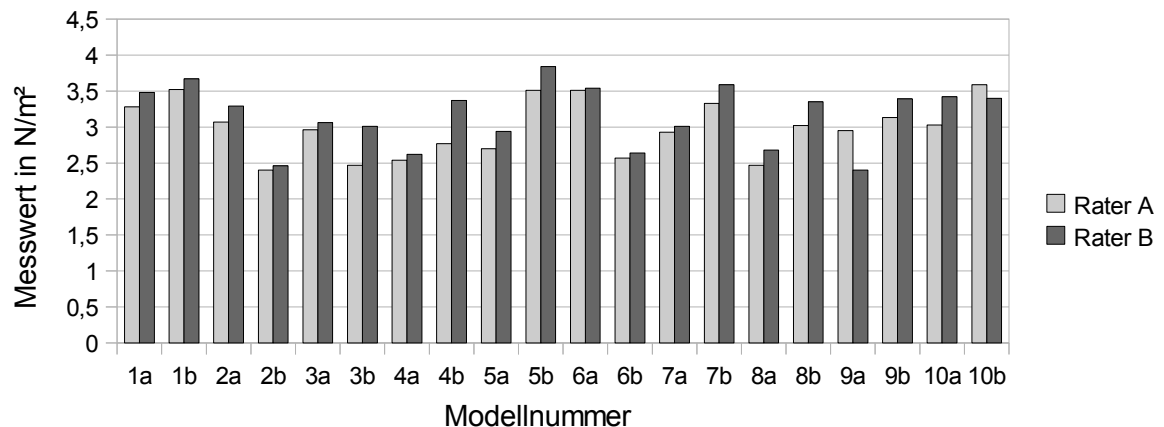


Diagramm 1: Mittelwertvergleich der Modelle mit „TTM“ zwischen den Ratern über alle MZP (n=60; entspricht pro Rater 5 Messungen bei 12 Punkten je Modell)

Es bestehen Unterschiede in den Mittelwerten zwischen den Modellen, demnach scheint das „TTM“ zwischen verschiedenen Modellen differenzieren zu können. In vielen Fällen messen die Rater beim gleichen Modell ähnlich hohe Werte. Allerdings sind auch teilweise stärkere Abweichungen zwischen den Ratern zu sehen, wie bspw. bei den Modellen 3b, 4b und 9a. Hier beträgt der absolute Messwertunterschied zwischen den Ratern 0,54 (Modell 3b), 0,55 (Modell 9a) und 0,6 (Modell 4b).

Das folgende Diagramm 2 zeigt die Minimal-, Mittel- und Maximalwerte die pro MZP jeweils von Rater A und Rater B mit dem „TTM“ gemessen wurden.

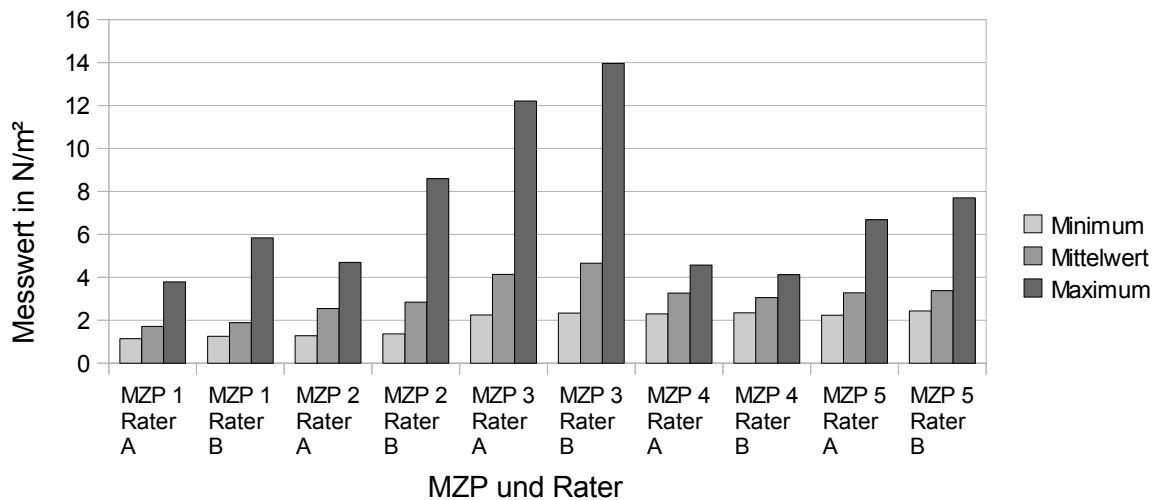


Diagramm 2: Minimum, Mittelwert und Maximum des „TTM“ pro Rater und pro MZIP (n=240)

Über die verschiedenen Messzeitpunkte verändern sich Minimum, Mittelwert und Maximum der Gesamtheit der Modelle. Zunächst ist auch hier festzustellen, dass Rater A niedrigere Werte gemessen hat als Rater B, außer zu MZIP 4. Die Minimalwerte weisen geringe Schwankungen auf, während die Maximalwerte stark variieren. Zu MZIP 3 wurden die höchsten Werte gemessen mit Maximalwerten von 12,21 (Rater A) und 13,95 (Rater B). Ab MZIP 3 steigen auch die gemessenen Minimalwerte an. Zuvor zwischen 1,14 und 1,37 gelegen, variieren sie ab MZIP 3 zwischen 2,23 und 2,43.

Der Mittelwert über alle Modelle steigt von MZIP 1 (RA 1,72; RB 1,88) zu MZIP2 (RA 2,54; RB 2,84) und MZIP 3 (RA 4,13; RB 4,65) kontinuierlich an und nimmt zu MZIP 4 (RA 3,27; RB 3,05) und MZIP 5 (RA 3,28; RB 3,38) wieder ab. Letztere MZIP haben ein ähnliches Mittelwertniveau. Eine ausgeglichene Verteilung von Minimum, Mittelwert und Maximum zwischen beiden Ratern scheint zu MZIP4 zu bestehen. Zu den restlichen MZIP gab es vermehrt Ausreißer in den Messwerten, sodass hohe Maximalwerte resultierten.

Die Modelle wurden hinsichtlich zweier Charakteristika eingeteilt: die Dicke ihrer Deckmatte und das Vorhandensein eines Holzbrettes. Dadurch entstehen vier Gruppen, deren Mittelwerte Rückschlüsse auf die Diskriminierungspotenz des Gerätes zulassen. In Tabelle 5 sind die Mittelwerte abhängig von den Gruppen dargestellt.

Tabelle 5: Gruppenunterschied „TTM“ pro Rater über alle MZP (pro Rater n=1200)

Mattendicke in mm	Holzbrett	Mittelwerte Rater A in N/m ²	Mittelwerte Rater B in N/m ²
8	ja	3,45	3,60
8	nein	3,32	3,48
5	ja	2,83	3,00
5	nein	2,60	2,79

Zunächst fällt auch hier auf, dass Rater A durchgängig niedrigere Werte maß als Rater B. Weiterhin verringern sich die Werte beider Rater vom Anfang zum Ende der Tabelle. Das Fehlen eines Holzbrettes sowie eine geringere Mattendicke führen in allen tabellarisch dargestellten Fällen zu niedrigeren Werten. Somit ist festzustellen, dass die Beschaffenheit der Modelle einen nachweisbaren Einfluss auf die Messwerte hat.

Zur statistischen Auswertung wurden für jeden Rater und pro Messpunkt die Werte über die fünf MZP gemittelt. Diese Mittelwerte wurden dann mittels ANOVA auf Unterschiede zwischen den Dicken der Deckmatten, dem Vorhandensein eines Holzbrettes und den Ratern untersucht. Die Varianzanalyse bestätigte bei einem Signifikanzniveau von 5% ($\alpha=0,05$) einen signifikanten Unterschied ($p<0,001$) der mittleren Messwerte zwischen den Modellen mit verschiedenen Dicken, zwischen den Modellen mit oder ohne Holzbrett, aber auch zwischen den Ratern. Somit sind diesbezüglich alle drei Arbeitshypothesen anzunehmen.

Das Betrachten von Häufigkeitsdiagrammen lässt den Schluss zu, dass das Gerät die Charakteristik der Modelle wiedererkennt - also in der Lage ist, die Härteverteilung der Modelle abzubilden. Existieren mehr Modelle mit weicheren Eigenschaften, ist eine Verschiebung der Häufigkeitskurve nach links, also zu niedrigeren Werten, zu erwarten. Im Umkehrschluss käme bei vorwiegend härteren Eigenschaften eine Rechtsverschiebung der Kurve infrage. Sind die Modelleigenschaften ausgewogen, müsste die Kurve symmetrisch sein. Die Form der Kurve dürfte über alle MZP und bei beiden Ratern gleich bleiben, da sich die Modelle zwischendurch in ihren Eigenschaften nicht verändern. Zugunsten der Übersichtlichkeit werden an dieser Stelle nicht alle Häufigkeitsdiagramme dargestellt. Die vollständige Darstellung aller Häufigkeitsdiagramme findet sich in „Anhang 2.pdf“. Exemplarisch sind die Kurven von MZP 1 in Diagramm 3 zu sehen. Es ist bei

Rater A (linke Kurve) und Rater B (rechte Kurve) eine Linksverschiebung der Normalverteilungskurve zu erkennen. Beide Kurven weisen einen steilen Anstieg und einen flachen Abfall der Messwerthäufigkeiten auf.

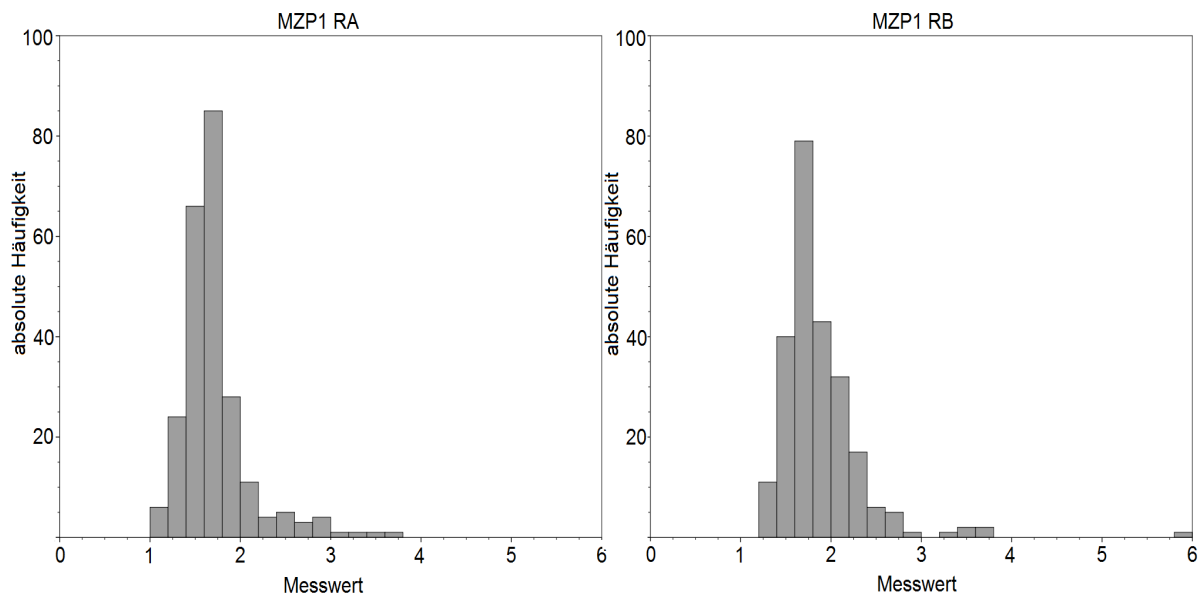


Diagramm 3: Verteilungskurve der Messwerte des „TTM“ zu pro Rater zu MZIP 1 (pro Diagramm n=240)

4.1.2 Intrasession-Reliabilität

Es wurden zehn Punkte zufällig bestimmt, indem zunächst die Modellnummer und anschließend die Nummer des Messpunktes auf dem jeweiligen Modell per Los gezogen wurden. Diese Punkte wurden zum ersten MZIP drei Mal hintereinander von Rater A gemessen (zum Studienablauf siehe Abbildung 14 Kapitel 3.4). Tabelle 6 zeigt die ICCs der Intrasession-Reliabilität sowie den p-Wert der Signifikanztestung des Unterschieds zum Testwert von 0,5.

Tabelle 6: ICC- und p-Werte der Intrasession-Reliabilität des „TTM“ von Rater A über alle und zwischen zwei Messungen (Overall n=30, zwischen zwei MZIP n=20)

	ICC	p
Overall	0,757	0,037
Messung 1+2	0,867	0,012
Messung 1+3	0,625	0,288
Messung 2+3	0,732	0,123

Das Gerät konnte eine zufriedenstellende Intrasession-Reliabilität von ICC=0,757 über alle drei Messwiederholungen erreichen. Die Alternativhypothese

$ICC_{\text{Intrasession}} \geq 0,5$ kann für den Overall-Wert und den Vergleich zwischen MZP 1+2 bestätigt werden: $ICC_{\text{Intrasession}}$ ist signifikant größer als 0,5. Für den Vergleich zwischen Messung 1+3 und Messung 2+3 muss die Nullhypothese angenommen werden.

4.1.3 Objektivität

Dieser Abschnitt soll die Frage klären, wie hoch die Objektivität des „TTM“ einzustufen ist. Im Verlauf zeigt sich eine hohe Varianz der Korrelation, dargestellt im folgenden Diagramm 4. Die höchsten Korrelationskoeffizienten wurden zu MZP 4 und MZP 5 erreicht. Bei MZP 3 scheint kein linearer Zusammenhang zwischen den gemessenen Werten zu bestehen ($r=0,094$).

Analog der Skala nach Bortz et al. 2008 (siehe Tabelle 3, Kapitel 3.5) ist die Korrelation zwischen den Ratern zu den ersten drei MZP nicht akzeptabel (MZP 1 $r=0,42$ und MZP 2 $r=0,426$). Ausreichende bis zufriedenstellende Ergebniskorrelation findet sich zu MZP 4 ($r=0,712$) und MZP 5 ($r=0,675$).

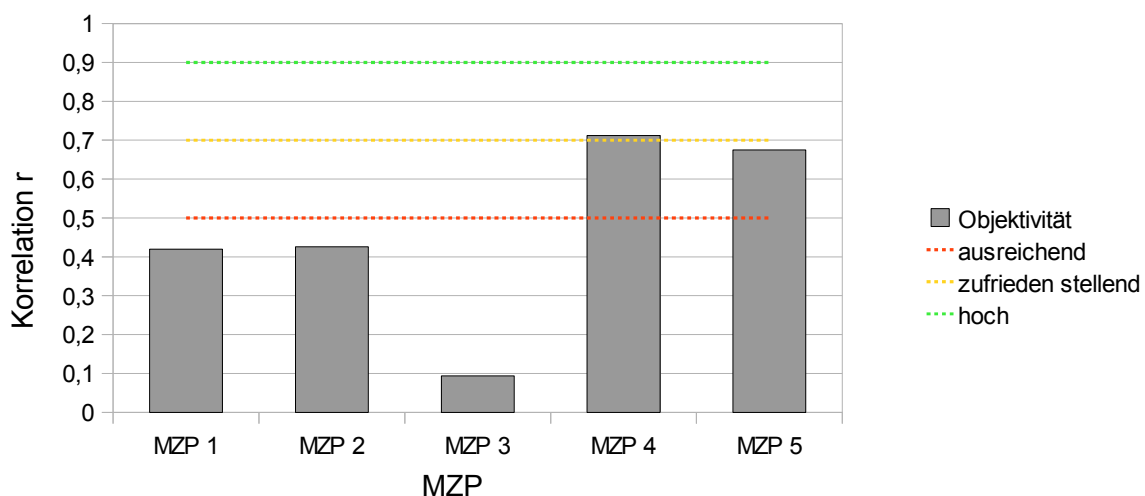


Diagramm 4: Objektivität (Korrelation r) des „TTM“ zwischen den Ratern pro MZP (jeweils n=480)

4.1.4 Test-Retest-Reliabilität

In diesem Abschnitt werden die Ergebnisse der Test-Retest-Reliabilität über fünf MZP und jeweils einen Rater dargestellt. Alle 240 Messpunkte (je zwölf Messpunkte an zwanzig Modellen) wurden für jeden MZP eines Raters untereinander verglichen.

Tabelle 7 zeigt die Intraklassen-Korrelation der einzelnen Rater über alle fünf MZP ($\text{Overall}_{\text{Einzel}}$) und im Vergleich zwischen jeweils zwei MZP.

Tabelle 7: Test-Retest-Reliabilität des „TTM“ pro Rater über alle und zwischen zwei MZP ($\text{Overall}_{\text{Einzel}}$ n=1200, zwischen zwei MZP n=480)

Messzeitpunkt	ICC Rater A	ICC Rater B
$\text{Overall}_{\text{Einzel}}$	0,126	0,080
1+2	0,138	0,158
1+3	0,001	0,009
1+4	0,188	0,109
1+5	0,101	0,070
2+3	-0,022 (0,000)	0,096
2+4	0,030	0,090
2+5	0,108	-0,028 (0,000)
3+4	0,125	0,085
3+5	0,291	0,087
4+5	0,575	0,382

Die Alternativhypothese $\text{ICC}_{\text{Retest}} \geq 0,5$ ist für alle ICCs abzulehnen, da der jeweilige p-Wert zur Signifikanztestung des Unterschieds zu 0,5 in nahezu allen Fällen 1,000 beträgt. Einzige Ausnahme bildet der Vergleich von MZP 4 mit MZP 5 von Rater A mit einem p-Wert von 0,051. Die ICCs der restlichen MZP reichen von -0,028 bis 0,382. Die negativen ICCs entstanden im Vergleich mit MZP 2. Besonders schwach ist auch die Korrelation von MZP 1 und MZP 3. Es heben sich, ähnlich den Ergebnissen der Objektivität, die Messzeitpunkte 4 und 5 positiv hervor.

Um den Verlauf der Korrelation nachvollziehbar zu gestalten, zeigt Diagramm 5 den ICC des einzelnen Raters über alle MZP ($\text{Overall}_{\text{Einzel}}$) sowie einen Vergleich der jeweils aufeinanderfolgenden MZP.

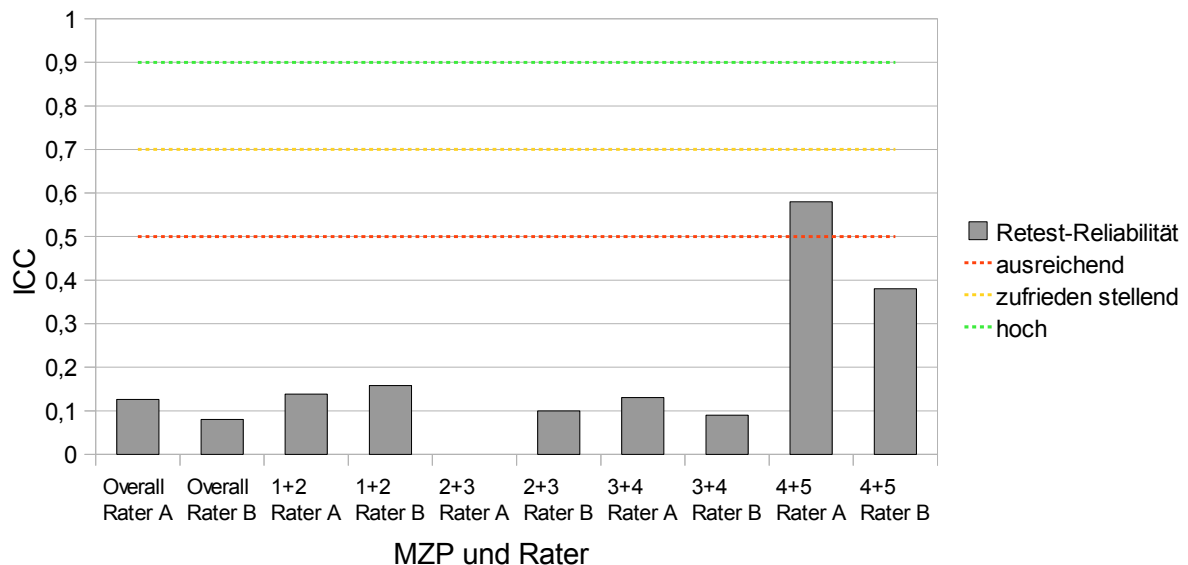


Diagramm 5: Test-Retest-Reliabilität des „TTM“ pro Rater über alle und zwischen den einzelnen MZP (Overall_{Einzel} n=1200, zwischen zwei MZP n=480)

Die Korrelationen der ersten drei MZP sind, wie bereits in Tabelle 7 dargestellt, sehr schwach. Es ist ersichtlich, dass MZP 1 und 2 minimal besser korrelieren als die nachfolgenden MZP 2 und 3 und MZP 3 und 4. Ähnlich den Objektivitätsergebnissen ist auch hier zunächst ein Abfall der Korrelation und gegen Ende ein steiler Aufstieg zu beobachten. Interessant ist ebenfalls, dass Rater A zu MZP 1 und 2 und MZP 2 und 3 eine schlechtere Korrelation aufweist als Rater B - danach jedoch besser abschneidet als Rater B.

Der Test-Retest-ICC für beide Rater gemeinsam über alle fünf Messzeitpunkte (Overall_{Gesamt}) beträgt **-0,044** und wird auf „Null“ gesetzt. Zur Signifikanzüberprüfung wurde der F-Wert ($F_{\text{intra}} = \mathbf{0,093}$) berechnet. Er liegt unter dem geforderten Grenzwert von 1,178.

Für die Overall_{Gesamt}-Berechnung, sowie ihre Prüfung auf Signifikanz, wurde ein anderer Rechenweg benutzt als für die Vergleiche zwischen den MZP eines Raters (siehe Kapitel 3.5). Zur besseren Übersicht werden die nach Eliasziw et al. 1994 berechneten Werte blau markiert. Auf diese Weise wird im gesamten Ergebnisteil verfahren.

Schließlich kann Rater A zu MZP 4 und 5 die Alternativhypothese bestätigen und

nach Bortz et al. 2008 eine ausreichende Reliabilität erreichen, für alle anderen MZP sowie die Overall-Berechnungen muss diese abgelehnt werden.

4.1.5 Interrater-Reliabilität

Für jeden MZP einzeln und über alle MZP zusammen wurden die jeweils 240 Messpunkte von Rater A mit denen von Rater B verglichen. Zu klären ist, ob die Interrater-Reliabilität mit einem ICC von mindestens 0,5 als ausreichend bezeichnet werden kann.

Die Ergebnisse sind ähnlich der Objektivität. An den ersten drei MZP ist die Korrelation zwischen den Ratern nicht akzeptabel. Zunächst erreicht sie zu MZP 1 und MZP 2 nahezu ausreichende Werte. Zu MZP 3 jedoch sinkt sie auf ein Minimum ab. Zufriedenstellende Ergebniskorrelation wird zu MZP 4 erreicht, etwas weniger zu MZP 5 (siehe Tabelle 8).

Der ICC der Overall-Berechnung ist kleiner als alle ICC-Werte der einzelnen MZP und weist keinen ausreichend hohen F-Wert ($F_{\text{inter}}=0,978$; Grenzwert 1,238) zur Ablehnung der Nullhypothese auf. Bei den ICCs der einzelnen MZP lässt sich für MZP 4 ($p<0,001$) und MZP 5 ($p=0,001$) die Alternativhypothese bestätigen, für MZP 1-3 ist sie abzulehnen ($p>0,900$). Dementsprechend ist die Reliabilität nach Bortz et al. 2008 für Overall und MZP 1-3 nicht ausreichend, zu MZP 4 zufrieden stellend und zu MZP 5 ausreichend.

Tabelle 8: Interrater-Reliabilität des „TTM“ zwischen den Ratern über alle und pro MZP (Overall $n=2400$, pro MZP $n=480$)

Messzeitpunkt	ICC
Overall	0,077
1	0,414
2	0,408
3	0,089
4	0,705
5	0,635

Zusammenfassend ließ sich in allen Teilabschnitten über die Ergebnisse des „TTM“ feststellen, dass die Werte stark abhängig von den MZP sind. Die Übereinstimmung zu MZP 1 und MZP 2 ist nicht ausreichend, zu MZP 3 am schwächsten ausgeprägt und zu MZP 4 und MZP 5 wieder ausreichend bis gut.

4.2 „Spineliner“

4.2.1 Mittelwertvergleiche

Der Vergleich der Mittelwerte der einzelnen Modelle über alle MZP (nachzulesen in Tabelle 9) ergibt Folgendes: Beim Parameter Widerstand misst Rater A bei allen Modellen höhere Werte als Rater B. Dieses Phänomen ist in geringerer Ausprägung auch bei Frequenz, Mobilität und Wellenform zu beobachten. Hier ist ebenfalls bei der Mehrzahl der Modelle der Mittelwert von Rater A höher als der von Rater B. Die Mittelwerte über alle Modelle bestätigen diese Tendenz. Weiterhin fällt auf, dass die Frequenz den geringsten Schwankungen innerhalb ihrer Messwerte unterliegt.

Tabelle 9: Mittelwerte des „Spineliners“ pro Rater über alle MZP (n=60 pro Modell, n=1200 über alle Modelle)

Modell	Wider- stand RA	Wider- stand RB	Frequenz RA	Frequenz RB	Mobilität RA	Mobilität RB	Wellen- form RA	Wellen-form RB
1a	41,0	35,6	38,6	38,3	56,5	54,8	52,0	50,7
1b	40,3	35,4	38,0	38,0	56,0	52,8	54,8	52,3
2a	35,8	31,9	38,1	37,7	59,4	56,3	54,5	51,7
2b	43,7	40,6	38,0	37,8	49,2	49,5	48,6	49,1
3a	44,5	37,2	38,4	37,8	51,3	51,2	50,3	50,1
3b	40,5	35,9	38,0	37,9	55,7	54,9	51,9	51,1
4a	46,2	40,3	38,4	37,9	54,1	52,9	51,9	50,8
4b	48,8	41,4	38,9	38,6	47,6	49,0	47,7	49,2
5a	45,0	39,0	36,8	36,7	51,9	52,4	52,5	52,6
5b	40,3	35,0	38,8	38,6	58,8	57,1	54,4	52,8
6a	35,5	31,5	37,5	37,5	55,5	51,7	51,9	48,5
6b	41,1	37,3	37,8	38,1	57,4	56,1	53,4	52,8
7a	39,3	35,7	38,6	38,6	58,7	56,6	53,5	52,4
7b	42,9	36,7	37,2	37,2	53,0	49,8	52,6	49,5
8a	41,1	34,6	38,3	38,5	56,3	58,2	52,0	53,2
8b	43,6	37,3	38,7	38,3	54,9	55,3	54,6	54,7
9a	42,9	39,1	37,4	37,7	50,5	50,0	49,9	49,0
9b	36,2	31,6	38,2	38,0	61,3	60,1	55,5	54,6
10a	41,5	38,6	38,8	38,1	46,9	46,1	46,4	46,0
10b	40,1	34,4	38,2	38,1	56,9	55,6	53,2	52,1
Alle Modelle	41,5	36,5	38,1	38,0	54,6	53,5	52,1	51,2

Zwischen den einzelnen Modellen sind Mittelwertunterschiede erkennbar, jedoch nicht immer deutlich. Das lässt sich durch grafische Darstellung der Mittelwerte der einzelnen Modelle über alle fünf MZP gut veranschaulichen. Diagramm 6 zeigt beispielhaft diese Werte für den Parameter Widerstand, da hier die Unterschiede zwischen den Ratern am deutlichsten werden. Es ist erneut sichtbar, dass Rater A in jedem Fall höhere Werte gemessen hat als Rater B, teilweise bis zu einem absoluten Mittelwertunterschied von 7,3 (Modell 3a) und 7,4 (Modell 4b). Die Diagramme der restlichen Parameter befinden sich in der Datei „Anhang 3.pdf“.

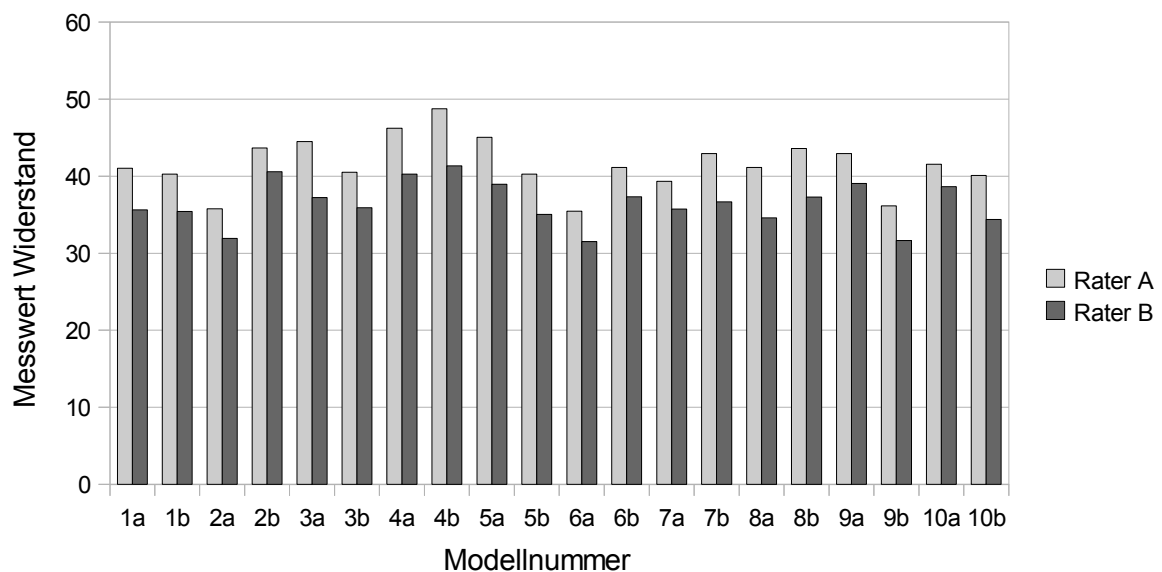


Diagramm 6: Mittelwerte des Widerstands der einzelnen Modelle pro Rater über alle MZP (n=60)

Die Parameter schwanken unterschiedlich stark in ihren Werten, wie anhand einer grafischen Darstellung der Minimal-, Mittel- und Maximalwerte eines jeden Parameters zu sehen ist (siehe „Anhang 4.pdf“). Neben der starken Variabilität ist ein deutliches Einknicken der Werte zu MZP3 (v.a. des Maximalwerts) auffällig. Besonders eindrucksvoll ist dies bei dem Parameter Mobilität zu sehen (Diagramm 7).

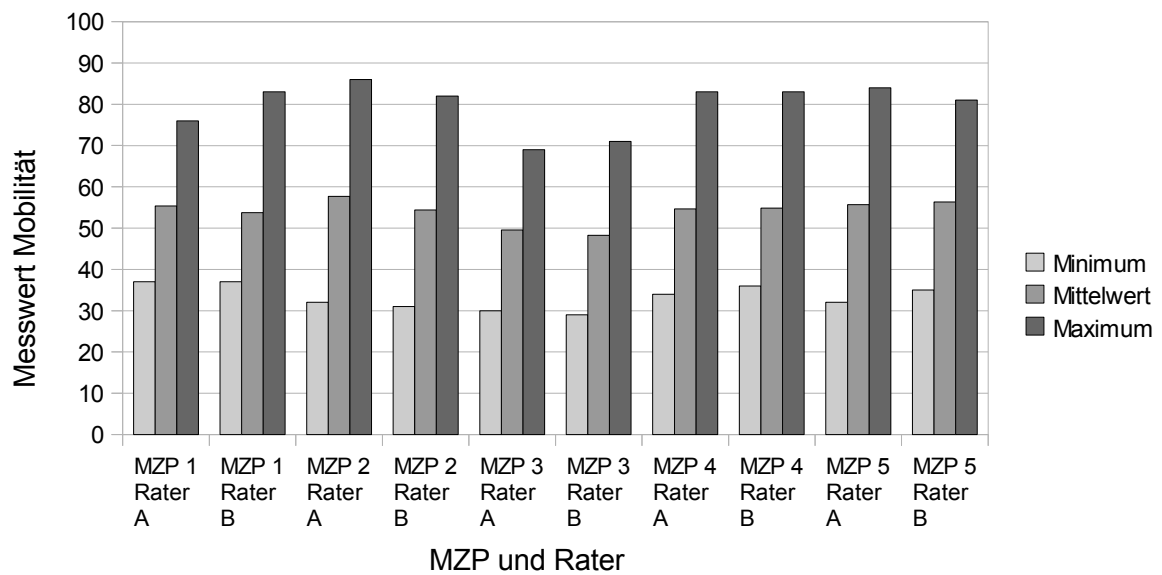


Diagramm 7: Minimal-, Mittel- und Maximalwerte der Mobilität pro Rater und pro MZIP (n=240)

Auch die anderen Parameter zeigen zu MZIP 3 ein unterschiedliches Verhalten: die Frequenz ist höher (abgebildet in Diagramm 8) und die Wellenform ist niedriger. Wellenform und Widerstand sind aus Platzgründen im Anhang zu vergleichen. Wie schon zuvor erwähnt, unterliegt die Frequenz den geringsten Schwankungen zwischen Minimal- und Maximalwert (Diagramm 8). Der Unterschied von Mobilität (Diagramm 7) und Frequenz (Diagramm 8) ist deutlich zu sehen.

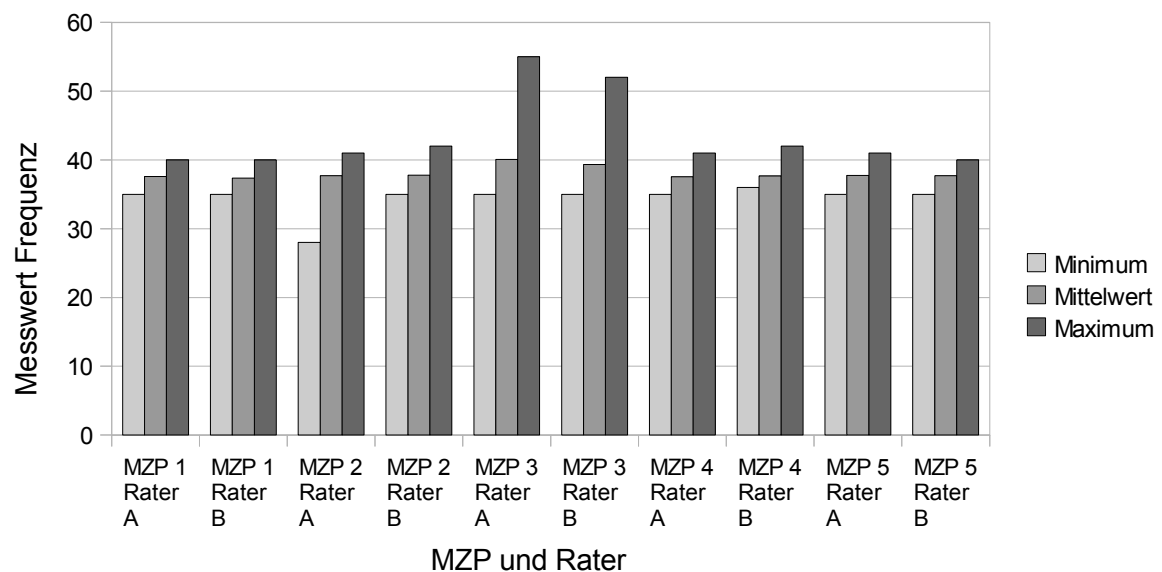


Diagramm 8: Minimal-, Mittel- und Maximalwerte der Frequenz pro Rater und pro MZIP (n=240)

Im Gruppenvergleich ist wieder erkennbar, dass Rater B meist niedrigere Werte maß als Rater A (siehe Tabelle 10).

Tabelle 10: Gruppenunterschied „Spineliner“ pro Rater über alle MZP (pro Rater und Parameter n=1200)

Mattendicke in mm	Holzbrett	Wider- stand RA	Wider- stand RB	Mobilität RA	Mobilität RB	Frequenz RA	Frequenz RB	Wellenform RA	Wellenform RB
8	ja	39,55	34,53	54,84	51,46	37,57	37,57	53,10	50,10
8	nein	38,65	33,71	58,54	56,78	38,36	38,17	53,90	52,37
5	ja	44,28	39,01	50,34	50,49	38,13	37,85	49,99	50,09
5	nein	41,66	36,76	56,41	55,73	38,20	38,17	52,52	52,05

Bei vorhandenem Holzbrett ist der Widerstand höher, Mobilität, Frequenz und Wellenform hingegen niedriger. Für jeden einzeln getesteten Parameter ist nach der Varianzanalyse mittels des ANOVA-Modells (siehe Kapitel 3.5) der Unterschied zwischen den Mattendicken, den Modellen mit und ohne Holzbrett sowie zwischen den Ratern signifikant (jeweils $p < 0,001$), die Alternativhypothesen können bestätigt werden. Das Gerät scheint Modellcharakteristika differenzieren zu können.

Diesen Schluss legen auch die Häufigkeitsdiagramme nahe, da sich die Kurven im Vergleich zwischen den Ratern und den MZP (Ausnahme MZP 3) stark ähneln. Aufgrund des großen Umfangs sind die Häufigkeitsverteilungen aller vier Parameter im Anhang zu finden. Der Widerstand, die Frequenz und die Wellenform zeigen eine nahezu symmetrische Häufigkeit der einzelnen Werte. Bei der Mobilität ist die Kurve leicht nach rechts verschoben, dennoch besteht zwischen den Ratern eine ähnliche Verteilung. In Diagramm 9 ist zum Vergleich MZP 1 von Rater A (links) und MZP 1 von Rater B (rechts) dargestellt.

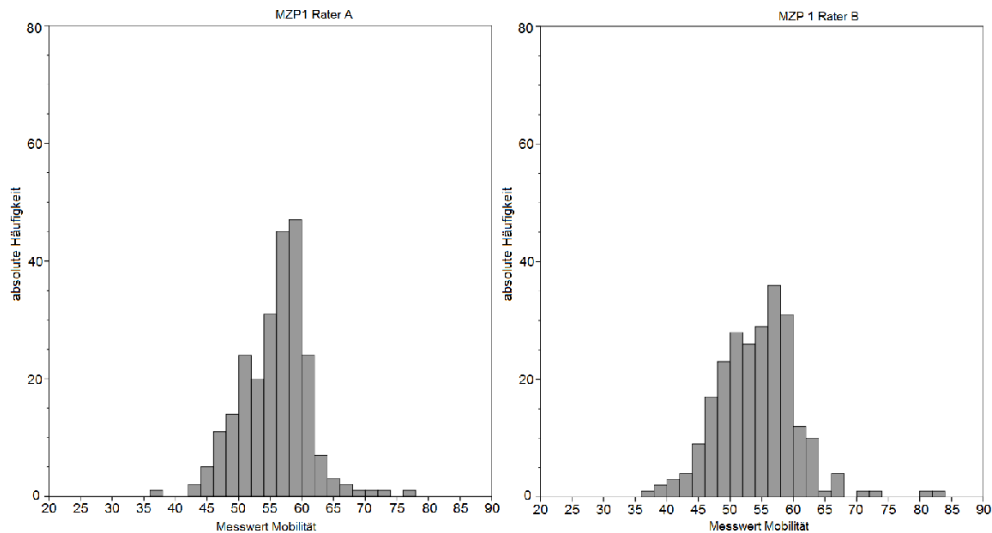


Diagramm 9: Beispiel Häufigkeitsverteilung der Mobilität pro Rater zu MZP 1
(pro Diagramm n=1200)

Zu MZP 3 finden sich Veränderungen in der Häufigkeitsverteilung der Messwerte: die Mobilitätskurve wird zweigipfelig und die Frequenzkurve (zuvor symmetrisch) läuft deutlich weiter nach rechts aus. Es wurden vermehrt höhere Frequenzwerte zu MZP 3 gemessen als zu den anderen MZP. Dadurch sind die Häufigkeiten auf eine größere Wertepalette verteilt. In Diagramm 10 sind zur Veranschaulichung MZP 3 und MZP 4 der Frequenz von Rater A dargestellt.

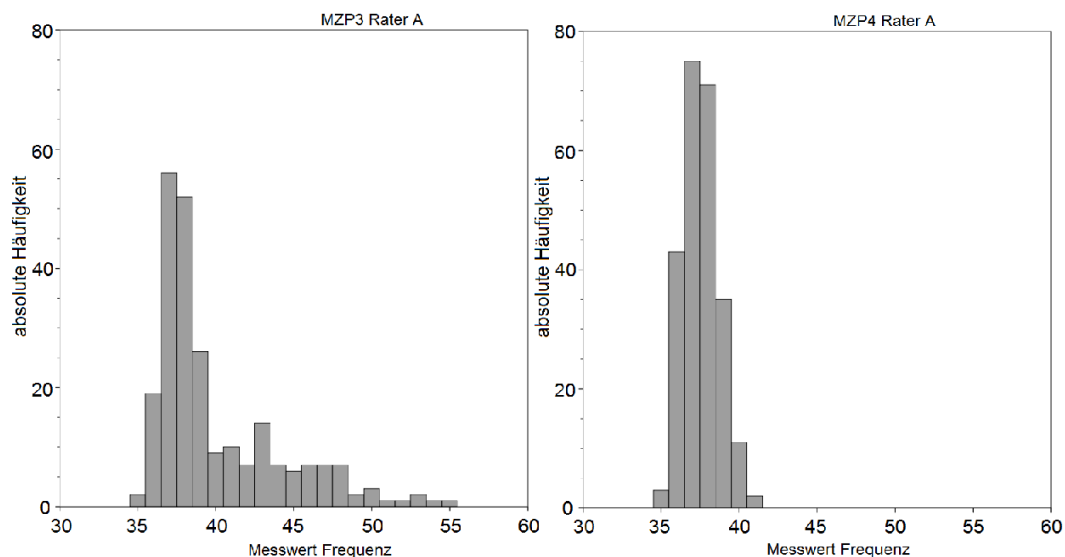


Diagramm 10: Beispiel Häufigkeitsverteilung der Frequenz von Rater A zu MZP 3 und 4
(pro Diagramm n=1200)

4.2.2 Intrasection-Reliabilität

Betrachtet man die Ergebnisse der Overall-Berechnung, erreichen alle Werte mindestens eine ausreichende Intrasection-Reliabilität, wobei die Frequenz die höchste Übereinstimmung (ICC 0,761) und die Wellenform die niedrigste Übereinstimmung (ICC 0,507) zeigt. Mobilität und Frequenz erreichen im Vergleich aller Messungen einzeln je eine mindestens ausreichende Reliabilität (ICC 0,529-0,813), während Widerstand (ICC 0,366-0,757) und Wellenform (0,259-0,744) in der Reliabilität stark schwanken (siehe Diagramm 11 und Tabelle 11).

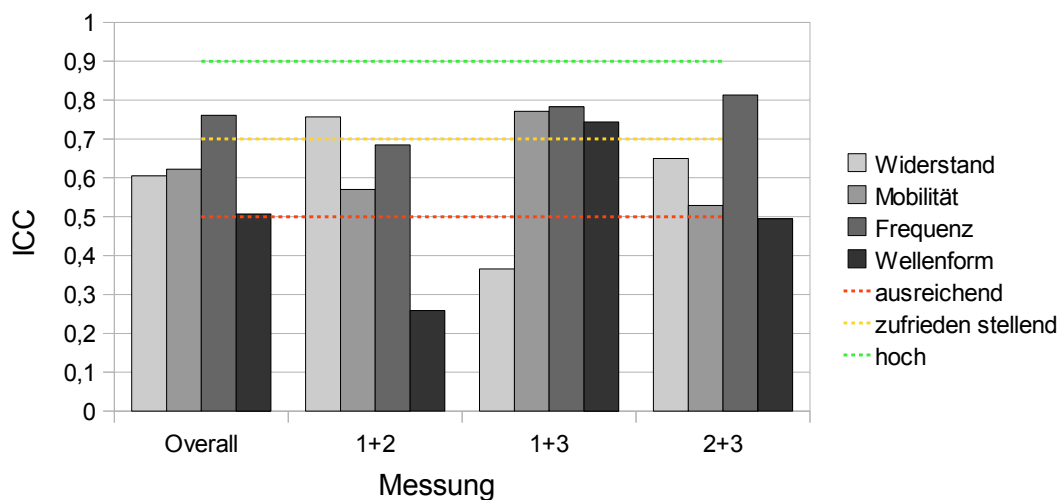


Diagramm 11: Intrasection-Reliabilität des „Spineliners“ von Rater A über alle und zwischen zwei Messungen (Overall n=30, zwischen zwei Messungen n=20)

Tabelle 11: Intrasection-Reliabilität des „Spineliners“ von Rater A über alle und zwischen zwei Messungen (Overall n=30, zwischen zwei Messungen n=20)

Messung	Widerstand	Mobilität	Frequenz	Wellenform
Overall	0,605	0,622	0,761	0,507
1+2	0,757	0,570	0,685	0,259
1+3	0,366	0,771	0,783	0,744
2+3	0,650	0,529	0,813	0,495

Die Alternativhypothese $ICC_{Intrasection} \geq 0,5$ lässt sich für die Frequenz Overall ($p=0,034$) und zwischen Messung 2 und 3 ($p=0,041$) bestätigen. Trotz der insgesamt zufriedenstellenden Werte reicht die Signifikanz für die anderen Parameter nicht aus, um die Nullhypothese abzulehnen.

Die Mobilität erreicht durchgehend ausreichende ICC-Werte über 0,5. Im Vergleich

zwischen zwei Messungen liegen Wellenform und Widerstand teilweise unter der Grenze von 0,5, teilweise aber auch im zufriedenstellenden Bereich. Auffällig ist, dass der Widerstand einen hohen ICC erreicht, wenn die Wellenform einen niedrigen ICC hat und umgekehrt.

4.2.3 Objektivität

Die Korrelation der Messwerte zwischen Rater A und Rater B reicht von 0,136 bis 0,865. Auffällig ist, dass alle Werte zu MZP 3 deutlich besser korrelieren als zu den anderen MZP. Bei allen MZP, außer MZP 3, besteht beim Parameter Frequenz die höchste lineare Korrelation. Eine Übersicht über alle Werte der Objektivität findet sich in Diagramm 12 und Tabelle 12.

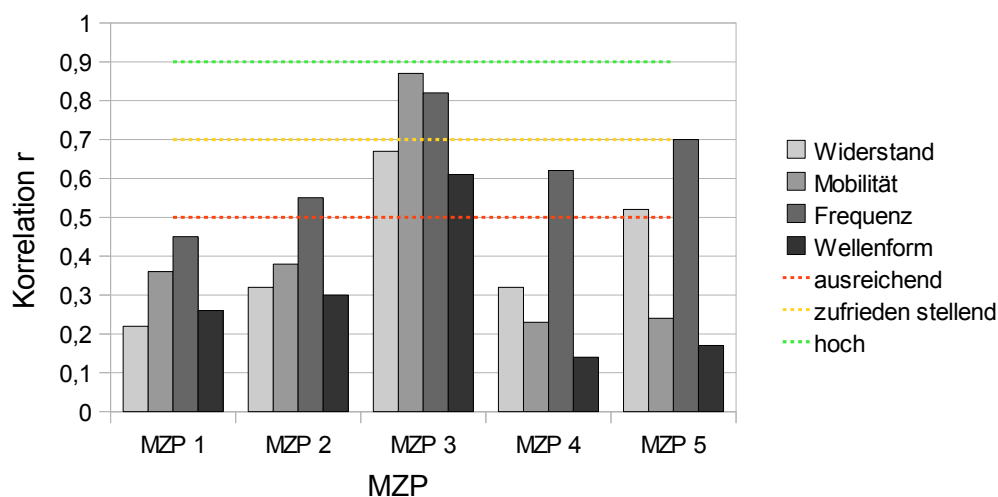


Diagramm 12: Objektivität (Korrelation r) aller vier Parameter des „Spineliners“ zwischen den Ratern pro MZP (n=480)

Tabelle 12: Objektivität (Korrelation r) aller vier Parameter des „Spineliners“ zwischen den Ratern pro MZP (n=480)

MZP	Korrelation Widerstand	Korrelation Mobilität	Korrelation Frequenz	Korrelation Wellenform
1	0,224	0,357	0,445	0,263
2	0,323	0,381	0,550	0,304
3	0,672	0,865	0,817	0,614
4	0,321	0,231	0,621	0,136
5	0,519	0,236	0,699	0,169

Zu MZP 3 erreichen alle Parameter ein ausreichendes Ergebnis. Mobilität und Frequenz weisen zu MZP 3 sogar ein zufriedenstellendes Ergebnis mit einem ICC von über 0,7 auf. Außerdem erreicht die Frequenz zu MZP 2 bis MZP 5 einen mindestens ausreichenden ICC. Der Widerstand überschreitet zu MZP 5 die Grenze von 0,5.

4.2.4 Test-Retest-Reliabilität

Die Ergebnisse der Test-Retest-Reliabilität fielen sehr heterogen und insgesamt schlecht aus. Eine mindestens ausreichende Reliabilität erreichten lediglich:

- Mobilität RA zu MZP 1+4 und MZP 4+5
- Frequenz RA zu MZP 1+2, MZP 1+4, MZP 1+5, MZP 2+4, MZP 2+5, MZP 4+5
- Frequenz RB zu MZP 1+5 und 2+5

Deutlich wird die insgesamt bessere Test-Retest-Reliabilität von Rater A gegenüber Rater B. Die Frequenz zeigt ebenfalls höhere Werte als die anderen Parameter, solange MZP 3 nicht involviert ist. Sobald MZP 3 in die Berechnung einbezogen wird, kommt es bei der Frequenz zu negativer Korrelation (siehe Tabelle 13, rot markiert). Dieses Phänomen passt zu der Feststellung, dass sich die Mittelwerte der Frequenz an MZP 3 deutlich von denen der anderen MZP unterscheiden (Kapitel 4.2.1).

Tabelle 13: Test-Retest-Reliabilität (ICC) des „Spineliners“ pro Rater über alle und zwischen zwei MZP (Overall_{Einzel} n=1200, zwischen zwei MZP n=480)

MZP	ICC Widerstand Rater A	ICC Widerstand Rater B	ICC Mobilität Rater A	ICC Mobilität Rater B	ICC Frequenz Rater A	ICC Frequenz Rater B	ICC Wellenform Rater A	ICC Wellenform Rater B
Overall _{Einzel}	0,275	0,255	0,289	0,217	-0,081 (0,000)	-0,065 (0,000)	0,252	0,169
1+2	0,381	0,390	0,381	0,437	0,607	0,481	0,306	0,368
1+3	0,141	0,157	0,261	0,220	-0,230 (0,000)	-0,180 (0,000)	0,012	0,033
1+4	0,393	0,225	0,518	0,298	0,621	0,392	0,455	0,222
1+5	0,373	0,242	0,371	0,337	0,725	0,539	0,309	0,338
2+3	0,122	0,275	0,170	0,322	-0,233 (0,000)	-0,252 (0,000)	0,008	0,084
2+4	0,423	0,354	0,466	0,302	0,581	0,460	0,397	0,231
2+5	0,374	0,333	0,383	0,138	0,647	0,638	0,319	0,077
3+4	0,203	0,241	0,285	0,211	-0,217 (0,000)	-0,198 (0,000)	0,091	0,050

MZP	ICC Widerstand Rater A	ICC Widerstand Rater B	ICC Mobilität Rater A	ICC Mobilität Rater B	ICC Frequenz Rater A	ICC Frequenz Rater B	ICC Wellenform Rater A	ICC Wellenform Rater B
					(0,000)	(0,000)		
3+5	0,140	0,113	0,203	0,031	-0,276 (0,000)	-0,254 (0,000)	0,019	-0,079 (0,000)
4+5	0,446	0,327	0,532	0,275	0,659	0,458	0,449	0,249

Nachfolgend dargestellt sind Grafiken der Test-Retest-Reliabilität aller benachbarten MZP einzeln im Ratervergleich sowie die Overall_{Einzel}-Test-Retest-Reliabilität von Rater A im Vergleich mit Rater B. Es bietet sich das gemeinsame Studium der Grafiken mit der obenstehenden Wertetabelle an.

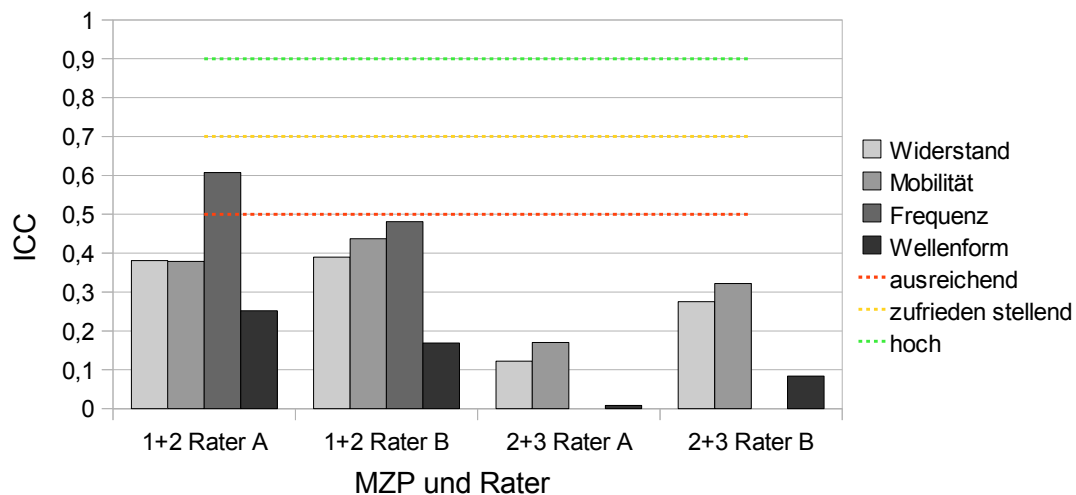


Diagramm 13: Test-Retest-Reliabilität des „Spineliners“ 1 pro Rater zwischen zwei MZP (n=480)

Diagramm 13 und 14 zeigen die Test-Retest-Reliabilität beider Rater im Vergleich der jeweils aufeinanderfolgenden MZP. Es fällt eine insgesamt schlechte Korrelation auf. Lediglich zu MZP 1+2 von Rater A und zu MZP 4+5 von Rater A überschreiten Frequenz bzw. Mobilität die Grenze von 0,5. In Diagramm 13 zeigt Rater B häufiger bessere Übereinstimmung als Rater A, in Diagramm 14 ist es umgekehrt. Alle Parameter weisen eine deutlich schlechtere Korrelation auf, wenn mit MZP 3 verglichen wird. Die Frequenz sticht in beiden Diagrammen heraus. Entweder erreicht sie die höchsten ICC-Werte oder sie nimmt negative Werte an, wenn MZP 3

involviert ist. Die Wellenform weist, neben den Ausreißern der Frequenz, immer die niedrigsten ICC-Werte auf. Der ICC der Mobilität ist, vor allem bei Rater A, oftmals höher als der des Widerstands.

Zu MZP 4+5 erreicht Rater A, zu MZP 1+2 erreicht Rater B die insgesamt beste Übereinstimmung innerhalb der Diagramme.

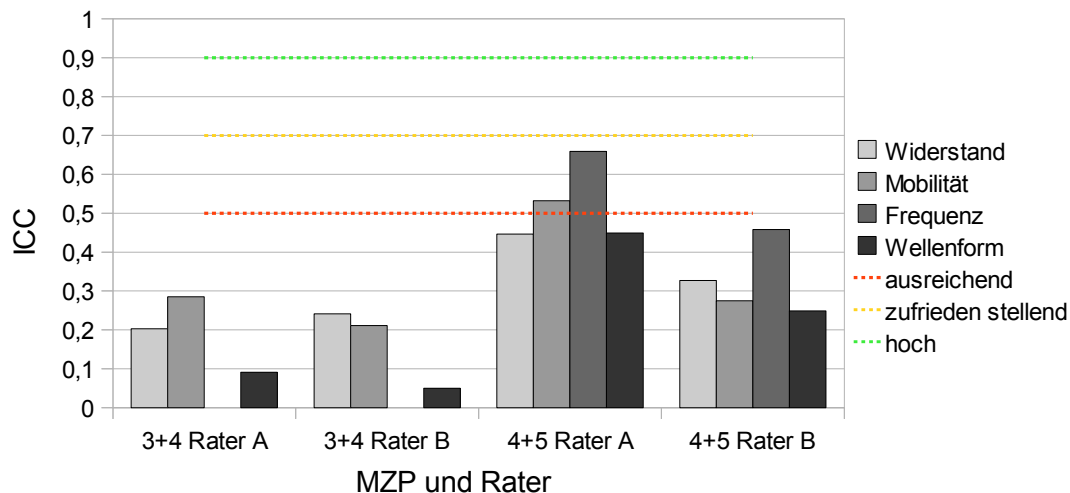


Diagramm 14: Test-Retest-Reliabilität des „Spineliners“ 2 pro Rater zwischen zwei MZP (n=480)

In der Overall_{Einzel}-Betrachtung fällt auf, dass alle Parameter unter einem ICC von 0,5 liegen (siehe Diagramm 15). Die Frequenz nimmt bei beiden Ratern einen negativen Wert (als „Null“ dargestellt) an. Rater A erreicht bei allen Parametern, außer der Frequenz, höhere Werte als Rater B. Der höchste ICC wird von Rater A bei der Mobilität erreicht, knapp dahinter rangiert der Widerstand. Bei Rater B ist es genau umgekehrt. Die Wellenform weist neben der Frequenz den niedrigsten ICC auf.

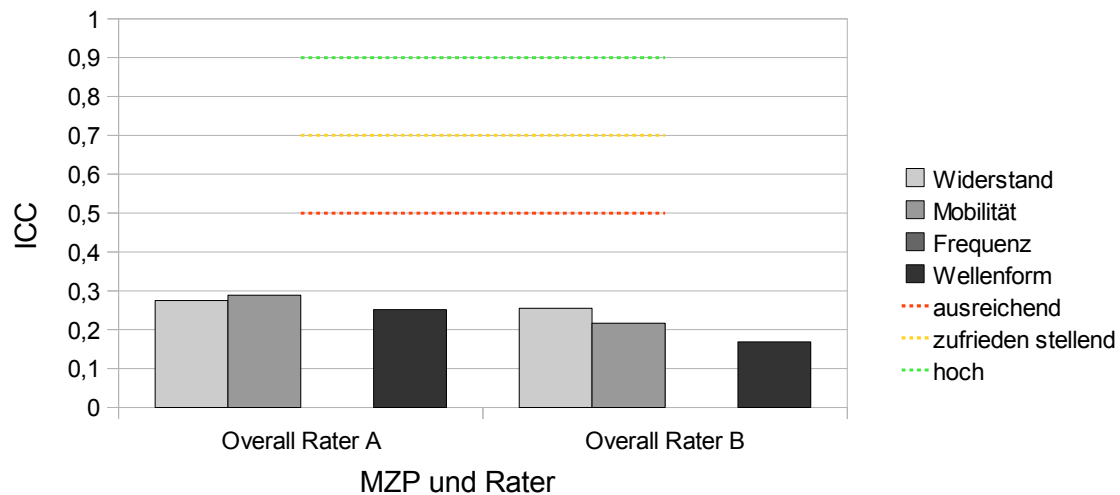


Diagramm 15: Test-Retest-Reliabilität des „Spineliners“ 3 pro Rater über alle MZP (n=1200)

Abschließend muss die Alternativhypothese $ICC_{Retest} \geq 0,5$ in der Overall_{Einzel}-Betrachtung für alle Parameter abgelehnt werden. Im Vergleich zwischen zwei MZP erreicht die Frequenz bei Rater A zu MZP 1+2 ($p=0,009$), MZP 1+4 ($p=0,003$), MZP 1+5 ($p<0,001$), MZP 2+4 ($p=0,038$), MZP 2+5 ($p<0,001$) und zu MZP 4+5 ($p<0,001$); bei Rater B zu MZP 2+5 ($p=0,001$) ein signifikant über 0,5 liegendes Ergebnis. Für die restlichen Parameter und MZP gilt die Nullhypothese.

Die Test-Retest-ICCs (Overall_{Gesamt}) betragen 0,354 für den Widerstand, 0,353 für die Mobilität, 0,071 für die Frequenz und 0,265 für die Wellenform. Bei allen Parametern liegt der F-Wert der Signifikanztestung unter dem Grenzwert von $F=1,178$. Somit wird die Alternativhypothese abgelehnt. Nach Bortz bedeutet dies eine nicht ausreichende Reliabilität, da die ICCs unter 0,5 liegen.

4.2.5 Interrater-Reliabilität

Analog zur Objektivitätsberechnung weist die Intraklassen-Korrelation der Messwerte zwischen Rater A und Rater B eine große Spannweite von 0,128 bis 0,856 auf. Die Frequenz ist mit Werten zwischen 0,445 und 0,802 am konstantesten in ihrer Reliabilität und erreicht als einziger Parameter, außer zu MZP 1, immer die Grenze von 0,5. Nachzulesen sind die ICC-Werte in Tabelle 14.

Tabelle 14: Interrater-Reliabilität des „Spineliners“ zwischen den Ratern über alle und pro MZP (Overall n=2400, pro MZP n=480)

Messzeitpunkt	Widerstand	Mobilität	Frequenz	Wellenform
Overall	0,475	0,502	0,776	0,294
1	0,223	0,351	0,445	0,259
2	0,323	0,381	0,547	0,304
3	0,670	0,865	0,802	0,613
4	0,321	0,224	0,621	0,128
5	0,517	0,235	0,696	0,169

In Diagramm 16 sind die Werte der Tabelle grafisch dargestellt. Auffällig ist auch hier die große Schwankungsbreite des ICC. Deutlich sichtbar ist, dass alle Werte zu MZP 3 besser korrelieren als zu den anderen MZP. Bei allen MZP, außer MZP 3, besteht beim Parameter Frequenz die höchste Übereinstimmung. Die Wellenform weist, außer zu MZP 1, immer die niedrigste Übereinstimmung auf.

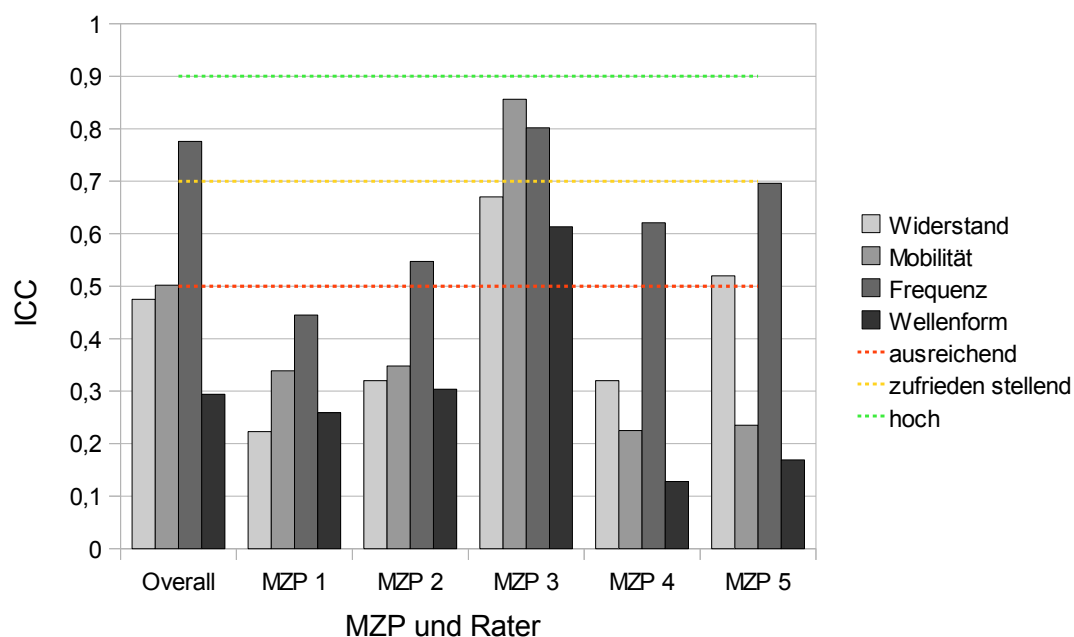


Diagramm 16: Interrater-Reliabilität des „Spineliners“ zwischen den Ratern über alle und pro MZP (Overall n=2400, pro MZP n=480)

Die Alternativhypothese $ICC_{\text{Interrater}} \geq 0,5$ kann zu MZP 3 für alle Parameter bestätigt werden ($p < 0,001$). Für die Frequenz gilt sie außerdem zu MZP 4 ($p = 0,003$) und MZP 5 ($p < 0,001$). In der Overall-Berechnung trifft sie für den Parameter Widerstand

($F_{\text{inter}}=1,897$), Frequenz ($F_{\text{inter}}=1,286$) und Mobilität ($F_{\text{inter}}=1,991$) zu, da diese Werte den Grenzwert von $F=1,238$ überschreiten.

Nach Bortz et al. 2008 ist die Interrater-Reliabilität aller Parameter zu MZP 3 und die der Frequenz zu allen Zeitpunkten außer zu MZP 1 ausreichend bis zufrieden stellend.

Zusammenfassend lässt sich feststellen, dass die Korrelation in Form der ICC-Werte stark abhängig von den MZP und den Parametern ist. MZP 3 sticht durch besonders gute Korrelation zwischen den Ratern hervor, gleichzeitig weist er jedoch eine schlechtere Korrelation im Vergleich mit den anderen MZP auf. Der Parameter Frequenz erreicht meist besonders hohe Übereinstimmungen, nimmt jedoch im Vergleich zwischen MZP 3 mit den restlichen MZP negative ICC-Werte an. Im Mittelwertvergleich wurden deutlich veränderte Frequenzwerte zu MZP 3 festgestellt. Die Wellenform ist oftmals der Parameter mit der niedrigsten Korrelation.

5 Diskussion

Ziel des folgenden Abschnitts ist, die Methoden und die Ergebnisse dieser Arbeit kritisch zu betrachten und mit aktueller Literatur zum Thema zu vergleichen. Die Vor- und Nachteile der genutzten Messgeräte sowie der Arbeit an Modellen sollen herausgearbeitet werden, sodass sich Empfehlungen für zukünftige Studien und die Anwendung in der Praxis ableiten lassen.

5.1 Ergebnisdiskussion

Die Ergebnisdiskussion soll eine sinnvolle Interpretation der Daten und Berechnungen liefern und sie in den aktuellen Forschungsstand einordnen. Der Abschnitt untergliedert sich in die Diskussion der Gütekriterien sowie die Diskussion der Ergebnisse des „Tissue Tensiometers“ und des „Spineliners“.

5.1.1 Gütekriterien

In der Manuellen Medizin (sowie in anderen medizinischen Fachbereichen) ist es verbreitet, ein Diagnose- oder Therapieverfahren auf Gütekriterien zu überprüfen. Wie bereits in Kapitel 1.1.2 beschrieben, wurden zahlreiche Studien und Literaturreviews zur Reliabilität von Palpationsuntersuchungen angefertigt, sowohl als Vergleich verschiedener Methoden als auch zwischen verschiedenen Untersuchern. In den meisten Fällen war die Reliabilität aufgrund unterschiedlichster Faktoren nicht zufriedenstellend. Im Zuge einer angestrebten Objektivierung der Untersuchung wurden vielfältige Geräte zur Diagnoseunterstützung entwickelt (siehe Kapitel 1.1.3).

Durch Reliabilitätsstudien können Erkenntnisse zum optimalen Anwendungsspektrum eines neuen Geräts gewonnen werden. Gut reproduzierbare Ergebnisse sind eine Voraussetzung dafür, dass das Gerät einen wichtigen Beitrag zur Diagnosefindung leisten kann. Variieren die Ergebnisse stark zwischen den Ratern, ist das Gerät nicht für die Nutzung durch wechselnde Untersucher zu empfehlen. Dann eignet es sich besser zur Verlaufskontrolle während einer Therapie.

Diese Arbeit leistet ihren Beitrag zum Forschungsstand der Geräte „Tissue Tensiometer“ und „Spineliner“ durch die Überprüfung von Objektivität und Reliabilität.

Keines der beiden Geräte wurde bisher auf diese Gütekriterien getestet. Die Reliabilitätsstudien zu verwandten Geräten fanden bisher ausschließlich an Probanden und mit geringen Fallzahlen statt (siehe Kapitel 1.3).

Objektivität entspricht der Unabhängigkeit eines Tests vom Untersucher (siehe Kapitel 1.2.1). Somit kann sie auch als Grad der Übereinstimmung zwischen den Ratern bezeichnet (Lienert und Raatz 1998) und durch die Pearson-Korrelation berechnet werden. Voraussetzung dafür sind numerische Daten, die von mehreren Ratern an vielen Datenträgern (z.B. Modellen) unabhängig voneinander gewonnen wurden (Amelang et al. 2006). Die Bedingungen treffen in dieser Arbeit zu: die Ergebnisse beider Messverfahren werden als Zahlen ausgegeben, es gab zwei Rater und insgesamt 240 verschiedene Messpunkte.

Eine Einteilung der Objektivität kann in drei Arten erfolgen, wie bereits in Abschnitt 1.2.1 beschrieben wurde. Die Durchführungsobjektivität erfordert einen standardisierten Aufbau und Ablauf einer Messung. In dieser Studie wurde eine hohe Standardisierung durch festgelegte Messpositionen und -abläufe (siehe Kapitel 3.4 und 5.2.2), sowie die Nutzung von unveränderlichen Modellen erreicht.

Auswertungsobjektivität ist durch die Geräte gegeben, indem sie automatisch die empfangenen Signale in Zahlen umwandeln. Der Prozess von der Messung bis zur Ausgabe der Messwerte läuft vom Rater unbeeinflusst innerhalb des Gerätes und der Software ab.

Die Interpretationsobjektivität konnte nicht untersucht werden, da hierfür ein Vergleich mit gültigen Referenzwerten notwendig ist. Da eigens angefertigte Modelle genutzt wurden, gab es keine Vergleichsmöglichkeit. Zudem wurden bisher für keines der beiden Geräte vom Hersteller Norm- oder Grenzwerte benannt. Eine Studie von Beyer et al. 2009 lieferte erste Referenzwerte für die Untersuchung mit dem Spineliner an gesunden, jugendlichen Probanden. Weitere Studien größeren Umfangs stehen aus, um Norm- und pathologische Grenzwerte an Probanden und Patienten zu ermitteln.

Für die Bewertung der Reliabilität wird nach Bühl 2008 üblicherweise der ICC-Wert berechnet, wie auch in einem Großteil der gefundenen Reliabilitätsstudien (Harrison et al. 2005, Leonard et al. 2003, Kachingwe und Phillips 2005, Ronchi et al. 2008, Taylor et al. 2010, uvm.). Dabei ist gegenüber dem Pearsonschen

Korrelationskoeffizienten (siehe Kapitel 3.5) von Vorteil, dass der ICC nicht nur die Richtung der Korrelation zwischen den Variablen, sondern auch das mittlere Niveau vergleicht (Bühl 2008).

Somit ist der ICC für diese Studie besser geeignet, da das Niveau der Messwerte zwischen den Ratern oder den MZP unbedingt übereinstimmen sollte, um eine klinisch relevante Aussage zur Reliabilität zu treffen. Der Anwender, in dem Fall der untersuchende Mediziner, soll die gemessenen Werte mit späteren Messungen oder denen anderer Ärzte vergleichen können. Um diese Vergleichbarkeit zu ermöglichen, muss das Gerät in der Lage sein, an Modellen ein konstantes Mittelwertniveau trotz verschiedener MZP oder Rater zu messen. Variieren die Werte bereits deutlich bei (unveränderlichen) Modellen, kann sich der Untersucher nicht auf ihre Genauigkeit verlassen. Daher kann nicht ausschließlich auf Richtung und Anstieg der korrelierenden Variablen geachtet werden. Die Nutzung des Pearson-Koeffizienten könnte in diesem Fall zu einer fehlerhaft hohen Reliabilität führen, da eine gleichgerichtete Korrelation mit einem anderen Mittelwertniveau gleichstark bewertet würde.

Bezüglich der Interpretation der ICC-Werte konnte innerhalb der Literaturrecherche zum Thema der Manualmedizin keine einheitliche Vorschrift gefunden werden. Verschiedene Einteilungen wurden entwickelt. Oftmals war nicht explizit beschrieben, welche Skala verwendet wurde (Assink et al. 2005, Lantz et al. 2003, Leach et al. 2003). Assink et. al 2005 gaben in ihrer Zielstellung einen ICC von über 0,75 an. Vermutlich wurde sich an Portney und Watkins 2000 orientiert. Diese bezeichnen Reliabilitäten mit einem ICC unter 0,5 als „schwach“, zwischen 0,5 und 0,75 als „moderat“ und über 0,75 als „gut“.

Kellis et al. 2008 verwiesen auf die Studie von Kachingwe und Phillips 2005. Letztere nutzten eine alte Unterteilung nach Meyers und Blesh 1962: unter 0,7 besteht „schwache“, von 0,7-0,79 „ausreichende“, von 0,8-0,89 „gute“ und über 0,9 „sehr gute“ Reliabilität.

Mehrere Studien verwendeten die Einteilung nach Shrout und Fleiss (Harrison et al. 2005, Dunk et al. 2004, Ronchi et al. 2008). Danach ist die Reliabilität bei einem ICC unter 0,4 als „schwach“, zwischen 0,4 und 0,75 als „ausreichend“ bis „gut“ und über 0,75 als „sehr gut“ anzusehen (Fleiss 1999).

Für diese Studie wurde die Skala nach Bortz genutzt, wie auch bspw. bei Bornmann

2009 und Vogel 2007. Der Unterschied zur Einteilung von Shrout und Fleiss besteht in der etwas kritischeren Betrachtung des ICC: ein „ausreichend“ wird erst ab 0,5 vergeben, ein „sehr gut“ ab 0,9 (Bortz 2008). Vergleichbar ist diese Skala mit der von Portney und Watkins. Ähnlich ist auch die Unterteilung nach Bühl 2008, sie unterscheidet jedoch noch zwischen geringer (0,21-0,5) und sehr geringer (0-0,2) Korrelation.

Für den Kliniker ist vermutlich die wichtigste Aussage, ab wann eine ausreichende Reliabilität besteht. Somit wäre die Grenze von 0,5 die Entscheidende. Danach wurden die Hypothesen in dieser Studie ausgelegt. Eine Einteilung der Reliabilität bei einem ICC unter 0,5 wurde nicht als sinnvoll erachtet, da ein Gerät mit einer derart niedrigen Reproduzierbarkeit aller Wahrscheinlichkeit nach keinen Beitrag zur höheren Objektivität einer manualmedizinischen Untersuchung leisten kann. Dazu müsste das Gerät mindestens ebenso zuverlässig messen wie ein Manualmediziner: in vielen Palpationsstudien konnte zumindest eine Reliabilität von $\kappa=0,5$ zwischen den Ratern nachgewiesen werden (Conradi und Smolenski 2005, Stochkendahl et al. 2006, Haneline und Young 2009). Der κ -Wert ist ebenfalls ein Reliabilitätskoeffizient und reicht wie der ICC von 0 bis 1.

5.1.2 „Tissue Tensiometer“

Der folgende Abschnitt widmet sich der Auswertung der Objektivitäts- und Reliabilitätsergebnisse des „TTM“.

Die Intrasession-Reliabilität liegt mit einem ICC von 0,76 im zufriedenstellenden Bereich (Bortz et al. 2008). Eine hohe Intrasession-Reliabilität ist eine gute Voraussetzung für eine ausreichende Objektivität, Test-Retest- und Interrater-Reliabilität. Da innerhalb einer „Session“ keine Veränderungen am Messaufbau, Messablauf, Rater, Gerät oder Modell vorgenommen werden, deutet eine niedrige Intrasession-Reliabilität auf generelle Probleme mit dem zu testenden Gerät oder dessen Handhabung hin. Dann empfiehlt es sich, eruierbare Fehler zu beheben, bevor die Testung der Gütekriterien fortgesetzt wird.

Im Anschluss begann die Messserie von fünf Terminen zur Überprüfung von Objektivität, Test-Retest- und Interrater-Reliabilität. Hier waren zwei Rater beteiligt, um den untersucherabhängigen Einfluss auf die Messergebnisse zu überprüfen. Wie

dem Ergebnisteil zu entnehmen ist, war die Objektivität und Interrater-Reliabilität an den ersten drei Terminen nicht ausreichend, danach im ausreichenden bis zufriedenstellenden Bereich. Eine mögliche Erklärung dieser Tatsache wäre ein stattgefundener Lerneffekt. Trotz einwöchiger Pause zwischen den Messungen schult der wiederholte Umgang mit dem Gerät den Untersucher in seiner Handhabung. Der Ablauf der Messung wird vertrauter und flüssiger. Dadurch können auch die gemessenen Daten zuverlässiger werden. Allerdings würde dann eine stetige Verbesserung der Korrelation von MZP 1 zu MZP 5 zu erwarten sein. Tatsächlich findet sich aber zu MZP 3 mit Abstand die schlechteste Korrelation im Vergleich zu den anderen Terminen.

Das Gerät bereitete zu MZP 1-3 Schwierigkeiten während der Messungen, worin eine weitere Ursache für die schlechte Korrelation an diesen Terminen liegen könnte. Nachdem etwa eine halbe Stunde gemessen wurde, kam es zu immer häufigeren Fehlermeldungen. Beispielsweise zeichnete das „TTM“ eine Kraft-Weg-Kurve auf, obwohl die Messung noch nicht gestartet wurde. Es gab häufig negative oder rückwärts verlaufende Kurven. Diese wurden nicht in die Auswertung einbezogen, sondern die Messung musste wiederholt werden. Da die Fehlerhäufigkeit im Laufe des Messtermins anstieg, nahm die Anzahl der Messwiederholungen zu. Bis zu fünfzig Wiederholungen waren teilweise pro Messpunkt nötig. Trotz der Selektion regelrecht verlaufender Kurven lieferten die letztlich gewerteten Kurven stark schwankende Werte innerhalb eines Modells.

Die Vermutung für diese hohe Fehlerrate ist eine Überlastung des Geräts. Zu den Terminen 1-3 war das „TTM“ schon zuvor für eine andere Studie in Benutzung gewesen. Außerdem stellten sich die Fehlermeldungen erst nach zunehmender Anzahl der gemessenen Modelle ein. Bei einer anderen, am Institut laufenden Studie mit weniger Messpunkten wurden diese Auffälligkeiten bislang nicht berichtet. Es könnte sein, dass das „TTM“ für eine hohe Messfrequenz (240 Messungen pro Rater) ungeeignet ist. Für die klinische Praxis ist eine derart hohe Messfrequenz nicht entscheidend. Vielmehr werden pro Patient wenige Punkte gemessen und die Zeitspanne zwischen den Messungen ist größer. In keinem Fall wird es notwendig sein, 480 Punkte direkt hintereinander zu messen.

Überlastung bzw. Überhitzung oder ein technischer Fehler des Geräts kann zu einer

sporadischen Fehlfunktion führen. Eine weitere Möglichkeit ist das Vorhandensein von elektromagnetischen Störwellen. Das Ausmaß an Störwellen durch Elektrosmog ist bisher wenig erforscht. Störwellen können von allen elektrischen Gegenständen ausgestrahlt werden. Dabei ist die Größenordnung der ausgestrahlten elektromagnetischen Felder nur schwer bestimmbar (Internationale Gesellschaft für Elektrosmog-Forschung 2011). Elektromagnetische Verträglichkeit (EMV) ist die Fähigkeit eines Geräts, während seiner Benutzung keine anderen Geräte zu beeinflussen.

Umgedreht ist die Störfestigkeit eines Geräts die Fähigkeit, während der Benutzung nicht durch elektromagnetische Felder beeinträchtigt zu werden. Die Leitlinie für EMV der Europäischen Union fordert für bestimmte Geräte, wie Medizinprodukte, eine CE-Kennzeichnung bevor sie auf dem europäischen Markt angeboten werden dürfen (Amtsblatt der Europäischen Union 2004). CE-gekennzeichnete Produkte versprechen unter anderem diese elektromagnetische Verträglichkeit, die Kennzeichnung wird vom Hersteller vergeben und nicht durch eine unabhängige Stelle (Horstkotte 2011).

Im Falle des „Tissue Tensiometers“ könnte eine mangelnde EMV zur Beeinträchtigung der Reliabilität geführt haben. Es ist möglich, dass zu den Messterminen 1-3 vermehrt elektromagnetische Störsignale zu verzeichnen waren. Diese könnten sowohl von Geräten innerhalb des Messraumes (Spiroergometer, EMG-Messgerät, Computer) oder angrenzenden Räumen des Instituts für Physiotherapie, als auch von anderen Abteilungen des Krankenhauses stammen (bspw. Radiologische Funktionsabteilung).

Als nächster Schritt bietet sich der Vergleich der Ergebnisse von verschiedenen Messzeitpunkten an. Lediglich zu einem Zeitpunkt erreicht die Test-Retest-Reliabilität den Zielwert ($ICC \geq 0,5$): bei Rater A zwischen MZP 4 und 5. Rater B weist ebenfalls zwischen diesen Terminen deutlich bessere ICC-Werte auf als zuvor ($ICC = 0,38$ im Vergleich zu $ICC < 0,16$). Diese Auffälligkeiten unterstreichen den zuvor geäußerten Verdacht einer Geräteüberlastung.

Der vergleichsweise deutlich schlechtere Overall_{Gesamt}-ICC sowohl der Intra- als auch der Interrater-Betrachtung, ist durch die Verwendung der Formel nach Eliasziw et al. 1994 zu erklären. Diese Formel bezieht alle fünf Messzeitpunkte und das Vorhandensein von zwei Ratern in die Beurteilung ein. Daher sorgen die

Messzeitpunkte mit geringer Korrelation (MZIP 1-3) für ein niedriges Gesamtergebnis.

Zudem ist die Variabilität eines Raters zwischen den verschiedenen Terminen größer, als die Variabilität zwischen den Ratern an einem Termin. Diese Tatsache unterscheidet sich von anderen Reliabilitätsstudien. In allen bisherigen Studien über das, dem „TTM“ ähnliche, „Myotonometer“ war die Test-Retest- bzw. Intrarater-Reliabilität höher als die Interrater-Reliabilität (Aarrestad et al. 2004, Leonard et al. 2003, Lidström et al. 2009). Ebenso verhält es sich mit den Studien anderer Geräte (Harrison et al. 2005, Kachingwe und Phillips 2005, Ronchi et al. 2008). Im Normalfall misst ein Rater auf eine bestimmte Art und Weise, die sich zwischen den Messungen nur geringfügig unterscheidet. Wird mit einem zweiten Rater verglichen, ist eine logische Konsequenz, dass mehr Schwankungen in den Werten zwischen beiden Ratern festzustellen sind, als in den Werten zwischen einem Rater.

Sind die Schwankungen zwischen einem Rater an zwei MZIP größer als zwischen zwei Ratern an einem MZIP, deutet das auf eine stärkere Abhängigkeit der Werte von dem MZIP als von dem Rater hin. Diese Tatsache lässt sich außer durch Geräteüberlastung nur schwer erklären. Theoretisch kann ein unterschiedlicher Messaufbau oder -ablauf ursächlich sein. Es wurde darauf geachtet, die Bedingungen konstant zu halten: die Unterlage für die Modelle, die Position des Raters, den Winkel in dem das Gerät aufgesetzt wird und die Messpunkte. Kleine Abweichungen von der Standardisierung sind einzukalkulieren, würden jedoch nicht die enormen Schwankungen des ICC erklären.

Bei Betrachtung der Rohwerte fällt auf, dass Rater A insgesamt niedrigere Werte gemessen hat als Rater B. Die nächstliegende Ursache ist eine geringfügig unterschiedliche Handhabung. Möglicherweise hat Rater A den Messkopf langsamer und mit weniger Kraft auf den Messpunkt gedrückt, sodass die Gewebespannung dementsprechend geringer ausfiel. Es wurde festgestellt, dass bei einer höheren voreingestellten Kraft höhere Gewebespannungswerte resultieren. Allerdings stoppt das Gerät die Messung automatisch bei der voreingestellten Kraft, sodass bei korrekter Funktion der Software zwischen Rater A und Rater B diesbezüglich keine Unterschiede bestanden haben dürften.

Für die Studie wurde entschieden, mit sechs N zu arbeiten, obwohl zwölf N die

Standardeinstellung war. Der Hersteller gab keine Hinweise zur Anwendung einer bestimmten Kraft, ebenso konnten in der Literatur keine Richtlinien gefunden werden. Die Vermutung war, je fester das Gewebe umso höher sollte die Voreinstellung sein. Da die Festigkeit zwischen den einzelnen Modellen variiert, wurde sich im Vorfeld für sechs N entschieden, da dieser Wert etwa in der Mitte des einstellbaren Newtonbereichs (von 1-15) liegt.

Leonard et al. 2003 berichten von einer besseren Reliabilität in den höheren Newtonbereichen. Allerdings wurde dort an Probanden und nicht am Modell gemessen. Zudem wurde das „Myotonometer“ und nicht das „Tissue Tensiometer“ genutzt, daher nur bedingt vergleichbar. Dennoch besteht die Möglichkeit, dass die Reliabilität mit einer Voreinstellung von zwölf oder mehr N besser ausgefallen wäre.

Eine weitere Voreinstellung des Herstellers ist die dreifache Messung eines Punktes. Daraus wird der Mittelwert bestimmt. Generell könnte dies zu einer höheren Genauigkeit der Ergebnisse führen, da sich zufällige Fehler bei steigender Anzahl der Messwiederholungen ausgleichen (Squires 1971). Dennoch wurde sich in dieser Studie gegen eine Messwiederholung entschieden, da mit dem „Spineliner“ nur eine Messung pro Punkt vorgesehen war und der Messablauf bei beiden Geräten angeglichen werden sollte. Aarrestad et al. 2004 nutzten in ihrer Studie über das „Myotonometer“ fünf Messwiederholungen und erreichten damit ICC-Werte zwischen 0,74 und 0,9. Allerdings hatten sie eine geringere Fallzahl (acht Messungen an zehn Probanden pro Rater) als diese Studie (zwölf Messungen an zwanzig Modellen pro Rater). Für die Anwendung am Patienten, mit wenigen Messpunkten pro Sitzung, kann eine Mittelung von mehreren Messungen pro Punkt aus o.g. Gründen sinnvoll sein.

Stellt man die Häufigkeiten der einzelnen Messwerte über alle Modelle in einer Grafik dar, zeigt sich eine Linksverschiebung der Normalverteilungskurve (siehe Diagramm 3 in Kapitel 4.1.1 und Anhang 2). Obwohl sich die Form der Kurven zwischen den MZP unterscheidet, bleibt die Linksverschiebung erhalten. Demzufolge wurden häufiger vergleichsweise niedrige Werte gemessen, also muss die Mehrzahl der Modelle eine geringere Festigkeit aufweisen. Tatsächlich existieren mehr Modelle mit einer 5mm dicken, weniger festen Deckmatte. Das „TTM“ scheint sie zu erkennen. Im Gegensatz dazu gibt es gleichviele Modelle mit bzw. ohne Holzbrett. Würde das

„TTM“ diesen Unterschied bevorzugt abbilden, wäre eine Normalverteilung zu erwarten. Daher lässt sich vermuten, dass die obere Matte für das „TTM“ eine größere Rolle spielt, als der Unterschied zwischen Modellen mit oder ohne Holzbrett. Diese Theorie wird durch den Mittelwertvergleich zwischen den Modellgruppen unterstützt (Werte siehe Tabelle 5 in Kapitel 4.1.1). Die Mittelwertunterschiede zwischen den Modellen mit unterschiedlicher Deckmatte sowie zwischen den Modellen mit oder ohne Holzbrett sind signifikant unterschiedlich, das Gerät kann also die Unterschiede differenzieren. Dennoch sind die Mittelwertunterschiede zwischen den verschiedenen Deckmatten deutlicher, als zwischen den Modellen mit und ohne Holzbrett. Anscheinend misst das Gerät also bevorzugt die oberen Schichten der Modelle. Mit dieser Möglichkeit wurde bereits während der Messung gerechnet, da der Stempel des „TTM“ kaum mehr als die obere Matte eindrückt. Als mögliche Schlussfolgerung lässt sich ziehen, dass das „TTM“ eher für oberflächliche Messungen geeignet ist. Da es für die Bestimmung des Muskeltonus konzipiert ist, der oberflächennah gemessen werden kann, hat diese Erkenntnis keinen Nachteil bezüglich der Eignung in der Praxis. Ein Beleg findet sich in der Literatur: in der bisher einzigen weiteren Studie über das „TTM“ schien die Praxistauglichkeit zufriedenstellend, immerhin 82% der palpatorisch identifizierten muskulären Triggerpunkte wiesen auch eine per „TTM“ gemessene erhöhte Gewebespannung auf (Buchmann et al. 2007).

Zusammenfassend kann das „TTM“ als Diagnoseunterstützung sowie für die Verlaufskontrolle einer Therapie empfohlen werden. Aufgrund der teilweise nicht zufriedenstellenden Objektivität und Reliabilität sollte es nicht als alleiniges diagnostisches Mittel verwendet werden, was auch nicht der Anspruch des Geräts ist. Dennoch erkennt das „TTM“ die Charakteristik der Modelle: es kann zwischen festen und weniger festen Strukturen unterscheiden. Der Arzt kann es als Hilfsmittel verwenden, um dem Patienten eine Visualisierung der Befunde und Therapieerfolge zu ermöglichen.

5.1.3 „Spineliner“

Im folgenden Kapitel werden die Messergebnisse und die Werte der Objektivität und Reliabilität des „Spineliners“ diskutiert.

Für eine statistische Auswertung eignen sich numerische Daten am besten. In der Praxis sind laut Hersteller jedoch die per Software ausgegebenen Kurven für die Untersuchung und Therapie mit dem „Spineliner“ repräsentativer. Um die Kurven in großem Umfang zu vergleichen, müssten sie ebenfalls in mathematisch erfassbare Parameter gegliedert werden, wie Beyer et al. 2009 bereits in ihrer Studie demonstrierten. Daher sollte durch das Einbeziehen aller vier, vom Gerät ausgegebenen, Variablen in die Auswertung ein Kompromiss gefunden werden. Ursprünglich war nur die Betrachtung der Mobilität geplant.

Analog zum „TTM“ wurde der „Spineliner“ an zehn Messpunkten in drei aufeinanderfolgenden Messungen auf Intrasession-Reliabilität geprüft. Über alle drei Messungen (overall) betrachtet erreichen alle vier Variablen einen ausreichenden ICC, maximal jedoch $ICC=0,76$. Trotz der insgesamt zufriedenstellenden Ergebnisse, kann bei der Intrasession-Reliabilität durchaus eine höhere Korrelation erwartet werden, da die Messungen direkt nacheinander erfolgen und somit jegliche Änderungen im Messaufbau ausgeschlossen werden können. Somit kann eine schlechter ausfallende Intrasession-Reliabilität nur durch andere, unbekannte Einflussfaktoren erklärt werden.

Cremonese et al. 1998 erhielten bei ihrer Intrasession-Testung des „S.M.A.R.T. Adjusters“, einem artverwandten Produkt des „Spineliners“, eine Genauigkeit von rund 99%. Dabei wurde auf einer Eichplatte und mithilfe einer automatischen Messvorrichtung getestet. In dieser Studie wurden mehrere, variable Modelle verwendet und von einem Untersucher gemessen. Somit wurde sich näher an der Realität orientiert, allerdings waren auch schlechtere Ergebnisse zu erwarten. Fraglich ist, ob die geringe Intrasession-Reliabilität damit ausreichend erklärt ist. Trotz der Vielfalt der Modelle waren sie innerhalb der Sitzung unveränderlich und die Messung lief standardisiert ab (siehe Kapitel 3.4.2).

Die Frequenz erreicht von allen vier Parametern die höchste Intrasession-Reliabilität, vermutlich da sie insgesamt das kleinste Wertespektrum aufweist. Diese Tatsache setzt sich zu den fünf MZP zur Feststellung der Objektivität, Test-Retest- und Interrater-Reliabilität fort. Im Mittelwertvergleich zwischen den Modellen unterscheidet sich die Frequenz nur geringfügig. Der Großteil der Frequenzwerte liegt zwischen 35 und 42, während die anderen Variablen ein deutlich breiteres

Spektrum bedienen.

Einzig zu MZP 3 nahm die Frequenz höhere Werte an als zu den anderen MZP. Objektivität und Interrater-Reliabilität aller Parameter zeigen an diesem Tag deutlich bessere Ergebnisse. Die Test-Retest-Reliabilität ist ebenfalls verändert. Sobald ein MZP separat mit MZP 3 verglichen wird, sieht man eine schlechtere Korrelation als zwischen den anderen MZP untereinander. Besonders deutlich wird dieses Phänomen bei der Frequenz, deren Korrelation negative ICC-Werte annimmt. Zu MZP 3 wurden im Vergleich mit den anderen MZP demnach deutlich unterschiedliche Werte gemessen. Die Rohwerte der Frequenz zu MZP 3 sind höher als zu den anderen MZP.

Erniedrigt sind zu MZP 3 Mobilität und Wellenform. Für ihre Berechnung spielt die Länge der Schwingung eine Rolle (die genaue Formel ist in Kapitel 3.2.2 zu finden, Beyer et al. 2009). Die Frequenz entspricht dem Kehrwert der Schwingungsdauer. Dadurch ist durch die Änderung der Frequenz eine Veränderung dieser Werte zu erklären. Der Widerstand hat sich nicht signifikant geändert.

Es lässt sich feststellen, dass die Frequenz zu MZP 3 die deutlichste Veränderung durchläuft. Die Werte Mobilität und Wellenform hängen mit der Frequenz zusammen, können also von dieser Betrachtung ausgelassen werden. Für eine Änderung der Schwingungsfrequenz zu MZP 3 kann ein organisatorisch bedingter Ortswechsel verantwortlich sein. Es wurde ein anderer Raumteil genutzt. Diese Änderung wurde zunächst nicht als entscheidend für die Studie betrachtet, da der Raumteil lediglich durch eine Schiebetür vom sonstigen Messraum abgegrenzt war. Der Untersuchungsaufbau (Liege, Position des Geräts, der Modelle und des Raters) blieb gleich. Dennoch zeigen die Ergebnisse, dass selbst kleinste Veränderungen der Messbedingungen einen starken Einfluss haben können. Für nachfolgende Studien und Anwendungen empfiehlt es sich daher, den Messort nicht zu wechseln oder bewusst die Auswirkung verschiedener Messorte zu überprüfen.

Wie bereits erwähnt, sind Objektivität und Interrater-Reliabilität zu MZP 3 deutlich erhöht, d.h. die gemessenen Rohwerte zwischen den Ratern stimmen besser überein. Es wäre also möglich dass zu MZP 3 eine Störgröße (bspw. Elektromog, siehe unten) entfallen ist. Umgekehrt könnte auch eine Störgröße hinzugekommen sein, die für eine kleinere Variabilität der Messwerte sorgt und damit

fälschlicherweise die Objektivität und Interrater-Reliabilität erhöht.

Eine zufällige Verbesserung der Korrelation ist bei einem so großen Unterschied unwahrscheinlich, zudem sich auch die Rohwerte verändert haben. Aus diesem Grund ist auch ein Lerneffekt eher auszuschließen. Weiterhin wäre dann eine stetige Verbesserung der Korrelation von MZP zu MZP zu erwarten (siehe Kapitel 5.1.2). Jedoch sinkt die Korrelation zu MZP 4 und MZP 5 wieder auf ein ähnliches Niveau wie zu MZP 1 und MZP 2.

Es muss also einen Effekt gegeben haben, der für die Messung einer höheren Schwingungsfrequenz sorgt und gleichzeitig eine Annäherung der Werte zwischen Rater A und Rater B bewirkt. Wie bereits in Kapitel 5.1.2 beschrieben, kommen hierfür elektromagnetische Störsignale infrage. Durch den Ortswechsel zu MZP 3 lässt sich eine unterschiedliche Intensität der Störwellen erklären. Im Raumteil des MZP 3 befanden sich weniger elektrische Geräte. Trotz der im Jahr 2004 erfolgten CE-Zertifizierung des „Spineliners“ (TTC 2008), die eine elektromagnetische Verträglichkeit garantieren soll (Horstkotte 2011), könnte der piezoelektrische Sensor des Geräts anfällig für bestimmte Störsignale sein und somit die Messungen der MZP 1, 2, 4 und 5 beeinträchtigt gewesen sein. Die Beeinflussbarkeit des „Spineliners“ von Störsignalen zu testen, bleibt nachfolgenden Studien überlassen.

Die Häufigkeitsdiagramme (siehe Kapitel 4.2.1 Diagramm 9 und 10, sowie Anhang 2) zeigen für die Parameter Widerstand, Frequenz und Wellenform eine annähernde Normalverteilung, während die Kurve der Mobilität nach rechts verschoben ist. Es scheint also etwas mehr Modelle mit einer höheren Mobilität zu geben. Diese Feststellung deckt sich mit den Ergebnissen des „TTM“, hier waren die Kurven nach links verschoben, also Richtung niedrigerer Gewebespannung. Eine niedrige Gewebespannung müsste mit einer hohen Beweglichkeit, also Mobilität, einhergehen. Die Charakteristik der Modelle spiegelt sich in der Häufigkeitsverteilung wieder, wird also durch den „Spineliner“ erkannt.

Zu MZP 3 sind die Häufigkeitskurven verändert: die Mobilität ist zweigipfelig und nach links erweitert, die Frequenz läuft weiter nach rechts aus. Dies zeigt, wie bereits geschildert, dass zu MZP 3 vermehrt hohe Frequenzen gemessen wurden. Daraus ergibt sich auch die niedrigere Mobilität, denn starre Gewebe schwingen auch laut

Hersteller mit höherer Frequenz (Rustler 2010).

Im Gegensatz dazu zeigt sich im Gruppenvergleich bei den Modellen mit Holzbrett eine etwas geringere Mobilität bei niedrigerer oder gleicher Frequenz. Die geringe Mobilität ist zu erwarten, da das Holzbrett ein Hindernis der Schwingung darstellen soll. Jedoch irritiert die niedrigere Frequenz, da sich die Variablen normalerweise gegenläufig verhalten sollen. Da die Gegenläufigkeit bei den Häufigkeitsdiagrammen korrekt dargestellt wurde, liegt die Vermutung nahe, dass die Zuordnung der Modelle in vier Gruppen (nach Deckmatte und Vorhandensein eines Holzbretts, siehe Kapitel 3.3) nicht wie erhofft funktioniert.

Der Vergleich der Mittelwerte der verschiedenen Gruppen (siehe Kapitel 4.2.1 Tabelle 10) zeigt deutliche Unterschiede zwischen Modellen mit bzw. ohne Holzbrett sowie mit 8 mm bzw. 5 mm dicker Deckmatte, die alle als signifikant getestet wurden. Die Deutlichkeit der Mittelwertunterschiede variiert zwischen den Parametern und Ratern - teils ist der Unterschied zwischen den verschiedenen Deckmatten größer, teils der Unterschied zwischen Modellen mit bzw. ohne Holzbrett.

Zum einen beweist es erneut, dass die Einteilung nach den zwei Hauptmerkmalen der Modelle keine einheitlichen Modellgruppen hervorbringt, die klar voneinander zu differenzieren sind. Eine ausführliche Diskussion der Modelle folgt in Kapitel 5.2.1. Zum anderen scheint das Holzbrett für den „Spineliner“ ebenfalls entscheidend zu sein, wohingegen beim „TTM“ vor allem die Deckmatten ausschlaggebend waren. Dadurch ist anzunehmen, dass der „Spineliner“ in tiefere Gewebeschichten vordringen kann, als das „TTM“. Dies ist ein wichtiger Vorteil für die Praxis, da der „Spineliner“ nicht nur Muskeltoni, sondern auch segmentale Dysfunktionen der Wirbelsäule identifizieren soll. Dafür muss er in der Lage sein, zwischen den Dornfortsätzen durch Haut-, Fett- und Muskelschichten bis auf die Ebene der Wirbelkörper zu messen.

Der „Spineliner“ erreicht bei nahezu keinem Parameter ausreichende Objektivität und Reliabilität und weist signifikante Messwertunterschiede zwischen den Ratern auf. Dennoch kann das Gerät zur Diagnoseunterstützung und Therapieverlaufskontrolle empfohlen werden, da es die Charakteristik der Modelle erkennen und wiedergeben konnte. Die Parameter sollten nicht einzeln, sondern im Zusammenhang

untereinander und im Vergleich mit der Messkurve betrachtet werden. Auf diese Weise kann das Gerät einen guten Beitrag zur visuellen Bestätigung der ärztlichen Diagnose und des Therapieerfolges leisten.

Die in dieser Studie festgestellte räumliche Abhängigkeit des „Spineliners“, durch stark abweichende Werte bei Messung in einem anderen Raumteil zu MZP 3, sollte in der Wahl des Messortes berücksichtigt werden.

5.2 Methodendiskussion

Das folgende Kapitel behandelt die Eignung der Modelle sowie den Aufbau und Ablauf der Studie. Dabei sollen mögliche Fehlerquellen diskutiert und Verbesserungsansätze geliefert werden.

5.2.1 Modelle

Um ein neues Gerät auf die Gütekriterien zu überprüfen, gibt es verschiedene Ansätze. Eine Möglichkeit ist, den Versuchsaufbau stark zu standardisieren, um Fehlerquellen und unerwünschte Einflüsse nahezu auszuschalten. Cremonese et al. 1998 testeten den „S.M.A.R.T. Adjuster“ an Eichplatten und mithilfe einer automatischen Messvorrichtung. Als Rater diente eine Konstruktion aus Hebel und Drucksensor, das Gerät war fest im rechten Winkel zur Messoberfläche eingespannt. Auf diese Weise wurde der menschliche Einfluss so gering wie möglich gehalten und Abweichungen der Messergebnisse können mit hoher Sicherheit Mängeln am Gerät zugeordnet werden. Mit dem „S.M.A.R.T. Adjuster“ konnte dabei eine sehr gute Reproduzierbarkeit erzielt werden.

Problematisch ist die Übertragbarkeit eines solchen Ergebnisses auf den klinischen Alltag. Die Bedienung durch einen Untersucher, die Anwendung an Patienten oder Probanden sowie die räumlichen Voraussetzungen einer Arztpraxis haben Einfluss auf die Varianz der Messwerte. Daher ist ein sinnvoller zweiter Schritt, einen Studienaufbau näher an der Realität zu entwerfen, ohne dabei komplett auf eine Standardisierung zu verzichten. Es bietet sich an, menschliche Rater an unveränderlichen Modellen zu nutzen. So kann beurteilt werden, ob das Gerät in der Praxis auch bei einem Wechsel des Untersuchers zuverlässige Werte liefert. Werden Rater und Messoberfläche variabel gewählt, kann nicht genau festgestellt werden, ob die Fehlerquelle seitens des Untersuchers oder der zu untersuchenden Oberfläche liegt. Diese Art von Zwischenschritt vor der Messung vom Menschen am Menschen

fand bisher nicht statt (siehe Kapitel 1.3).

Trotz der Unveränderlichkeit der Modelle sind sie für diese Studie so variabel wie möglich konstruiert worden, um realitätsnah zu verfahren. Durch die verschiedenen Größen und Formen der Modelle sowie unterschiedliche Kombinationen der Modellschichten konnte Variation erzielt werden (siehe Kapitel 3.3).

Die Einteilung der Modelle in vier Gruppen (siehe Kapitel 3.3) ermöglichte neue Fragestellungen bezüglich der Geräte, allerdings wurden auch die Nachteile der Vielfalt und zufälligen Zuordnung von Modelleigenschaften deutlich. Beide Geräte konnten signifikante Unterschiede zwischen den Modellgruppen nachweisen. Allerdings zeigten auch die Mittelwertvergleiche zwischen den Ratern signifikante Differenzen. Daher scheinen die Mittelwertunterschiede zwischen den Modellgruppen vorhanden zu sein, jedoch nicht deutlich genug, um den interindividuellen Messwertunterschied zu überbieten.

Die Matten, die die Muskel- und Fettschicht repräsentieren sollen, beeinflussen die Messungen anscheinend ebenfalls stark, sodass keine aussagekräftigen Gruppen entstehen. Es müsste exakt dieselbe Zusammensetzung bestehen um vergleichen zu können und selbst dann könnten die unterschiedlichen Größen und Formen der Modelle noch zu Messunterschieden führen. Eine gleiche Zusammensetzung der Schichten findet sich zudem nur bei wenigen Modellen. Diese geringe Zahl erlaubt keine Rückschlüsse auf einen Zusammenhang der Werte, da keine ausreichenden Vergleichsmöglichkeiten mit anderen Modellgruppen gegeben sind. Zudem ist nicht jede mögliche Kombination von Matten und Holzbrett durch ein Modell abgedeckt.

Die Unterschiede zwischen den Modellen waren, wie beabsichtigt, sehr groß. Im Nachhinein wäre möglicherweise eine kleinere Variabilität (gleiche Größe und Form der Kästen, nur zwei Matten) mit gleichmäßiger Verteilung der Modellzahlen pro Gruppe aussagekräftiger gewesen. Interessant wäre bspw. der Aspekt, welche Schichten von den Geräten in der Tiefe noch klar differenziert werden können, vor allem die Unterscheidung ob mit oder ohne Holzbrett. Es könnten dann die „Diagnosen“ der Geräte (Triggerpunkt bzw. segmentale Dysfunktion bei Modellen mit Holzbrett) mit der „Diagnose“ eines Arztes durch Palpation verglichen werden. Daraus könnten Erkenntnisse sowohl für die Diskriminationsfähigkeit der Geräte als auch über das Palpationsgeschick des Arztes gewonnen werden. Für solche Zwecke

sind Modelle sehr gut geeignet, da je nach Bedarf nur bestimmte Veränderungen an der Zusammensetzung vorgenommen werden können.

Nicholson et al. 2003 entschieden sich ebenfalls für Modelle zur Überprüfung des Einflusses der Viskosität auf die Palpationsergebnisse. Es wurde ein Gerät erschaffen, welches dem Untersucher automatisch verschiedene viskoelastische Stimuli präsentieren kann, wobei nur die Viskosität verändert wird. Fünfundzwanzig Untersucher mit unterschiedlichem Erfahrungsstand mussten fünf verschiedene Härtegrade unterscheiden, die je zwanzig Mal in zufälliger Reihenfolge präsentiert wurden. Dabei war zu beurteilen, ob der jeweils nachfolgende Stimulus härter oder weicher als der vorhergehende war. Durch die Nutzung eines Modells bzw. Geräts konnte eine Standardisierung der Härtegrade erreicht werden.

Die Zuverlässigkeit eines Geräts müsste an Modellen höher sein als an Probanden, da der Zustand der Modelle zwischen den einzelnen Messungen weitestgehend stabil bleibt. In dieser Studie konnten jedoch insgesamt keine besseren Ergebnisse erzielt werden, als in bekannten Studien verwandter Geräte an Probanden. Die ICCs der Intra- und Interrater-Reliabilitäten liegen bei Studien des „Myotonometers“ am Probanden zwischen 0,74 und 0,99 (Aarrestad et al. 2004), zwischen 0,75 und 0,99 (Leonard et al. 2003) oder zwischen 0,51 und 0,96 (Lidström et al. 2009). Probandenstudien mit „Spineliner“-ähnlichen Geräten erreichten ICCs zwischen 0,78 und 0,89 beim „PulStar“ (Leach et al. 2003) oder zwischen 0,67 und 0,86 beim „ProAdjuster“ (Tepe und Zhang, Jahr unbekannt). Für die niedrigeren Reliabilitäten in dieser Studie sind vermutlich die in der Ergebnisdiskussion (Kapitel 5.1) genannten Faktoren sowie ein unterschiedlicher Studienaufbau (siehe Kapitel 5.2.2) verantwortlich.

Als modellbezogene Ursache kann der Härtegrad der Modelle diskutiert werden. Möglicherweise messen die Geräte zuverlässiger am Menschen, weil sie nur für bestimmte Gewebespannungen bzw. Schwingungsfrequenzen geeignet sind. Dagegen spricht jedoch, zumindest beim „Spineliner“, die sehr gute Reproduzierbarkeit des geräteverwandten „S.M.A.R.T. Adjusters“ an der Kalibrierplatte (Cremonese et al. 1998). Es wird geschlussfolgert, dass die insgesamt unzureichenden Reliabilitäten nicht auf eine fehlerhafte Modellentwicklung zurückzuführen sind. In den vorhergehenden Kapiteln wurde bereits auf andere mögliche Ursachen eingegangen.

5.2.2 Ablauf und Aufbau der Studie

Diese Studie wurde als prospektive randomisierte Reliabilitätsstudie an Modellen geplant. Eine solche Studie hat es bisher weder zum „Tissue Tensiometer“ noch zum „Spineliner“ oder ihren verwandten Geräten gegeben (siehe Kapitel 1.3). Die vorhandenen Reliabilitätsstudien wurden mit vergleichsweise niedrigen Fallzahlen (ein bis 35 Probanden) angefertigt (Aarrestad et al. 2004, Cremonese et al. 1998, Leach et al. 2003, Leonard et al. 2003, Lidström et al. 2009).

Eine Pilotstudie von Tepe und Zhang an 64 Probanden wurde gefunden, jedoch ohne Randomisation. Bei Aarrestad et al. 2004 wurde die Reihenfolge der Rater randomisiert. Eine randomisierte kontrollierte Studie entspricht dem höchsten Evidenzgrad (Selbmann und Meisner 2000). In dieser Studie wurde die Reihenfolge des zuerst messenden Raters und des zu messenden Modells randomisiert. Dadurch können systemische Fehler, die eventuell zu Beginn des Messtermins und gegen Ende aufgrund sinkender Konzentrationsfähigkeit auftreten, auf alle Modelle und Rater gleichmäßig verteilt werden.

Um die Gefahr eines unsicheren und fehlerbehafteten Messablaufs zu Beginn jedes Termins gering zu halten, wurden zunächst je drei Modelle á zwölf Punkte von jedem Rater gemessen, ohne in die Wertung einzugehen. Kachingwe und Phillips 2005 nutzten ebenfalls den Vorteil einer solchen Probemessung. Somit konnte vor der eigentlichen Messung die Vorgehensweise rekapituliert werden. Diese war zuvor in einer Trainingsphase vereinbart worden. In dieser Phase konnte das Messen mit dem jeweiligen Gerät geübt und etwaige Fehlerquellen durch einheitlichen Umgang beseitigt werden, bspw. die Übergabe des „Spineliners“ um ein Verdrehen der Tips zu verhindern. Die Handhabung zwischen den Ratern sollte bestmöglich übereinstimmen. Solche Trainingsphasen nutzten auch Kachingwe und Phillips 2005 und Kellis et al. 2008.

Letztere planten ebenfalls zwischen den Messungen eine Pause von einer Woche. Für diese Studie war eine Pause wichtig, um eine zu starke Gewöhnung der Rater an die Geräte zu verhindern und somit eine stetige Verbesserung der Reliabilitäten durch Lerneffekte möglichst auszuschließen. Letztenendes sollte auch ein ungeübter Untersucher mit dem Gerät gute Ergebnisse erzielen können, nachdem sich mit der Bedienung vertraut gemacht wurde.

Es wurden fünf Messtermine vereinbart. Diese Anzahl übersteigt die Wiederholungshäufigkeit der anderen gefundenen Gerätestudien, die meist nur zwei Termine eingeplant hatten (Harrison et al. 2005, Kellis et al. 2008, Leach et al. 2003, Leonard et al. 2003, uvm.). Nach Shrout und Fleiss 1979 sind häufigere Wiederholungen nützlich, um die Zufallsvarianz des ICC zu senken. Dabei ist vor allem eine Verbesserung des ICC bei zwei Wiederholungen gegenüber einer Wiederholung zu verzeichnen. Zwei bis drei Messungen sind näher an der klinischen Realität und daher öfter in der Literatur vorzufinden, da auch an Patienten meist nur eine Eingangs- und Ausgangsuntersuchung mit eventuell einer Zwischenkontrolle stattfindet. Die festgelegte Anzahl von fünf erwies sich jedoch als glücklich: das „Tissue Tensiometer“ verzeichnete erst nach dem dritten Termin akzeptable Ergebnisse. Beim „Spinelineer“ schien bis zum dritten Termin ein Lerneffekt stattgefunden zu haben, der sich durch Termin vier und fünf wieder relativierte. Im Endeffekt ergab sich durch die fünf Termine eine höhere Aussagekraft.

Die Objektivität wurde parallel zur Reliabilität für alle Termine berechnet. Ursprünglich sollte sie nur an dem ersten Tag bestimmt werden und wegweisend für die weiteren Untersuchungen sein. Da sie jedoch unerwartet niedrig ausfiel, wurde beschlossen, den Verlauf der Objektivität weiter zu beobachten. Aufgrund der hohen Intrasession-Reliabilität wurde eine Verbesserung erwartet.

Wie bereits ausführlich geschildert, wurde auf einen standardisierten und reproduzierbaren Messaufbau geachtet. Die Position der Modelle und der Rater in Bezug zur Untersuchungsliege wurde festgelegt und eingehalten. Einmalig gab es einen organisatorisch begründeten Raumwechsel, der trotz gleicher Liege und Messaufbau zu erheblichen Messwertunterschieden führte (siehe Kapitel 5.1.3). Das Vorhandensein weiterer Geräte, z.B. ein Spiroergometer oder EMG-Messgerät, im Untersuchungsraum ließ den Einfluss elektromagnetischer Störwellen vermuten.

Die Messpunkte selbst wurden genaustens markiert und nummeriert. Somit wurde sichergestellt, dass zu jedem Zeitpunkt der exakt selbe Punkt gemessen wird. Andernfalls kann die Fehlerrate einer Messung steigen (Haneline und Young 2009). Eine Markierung verwendeten auch Leonard et al. 2003 und Aarrestad et al. 2004. Cremonese et al. 1998 verzichteten auf Markierungen und erhielten dennoch sehr

gute Ergebnisse. Allerdings dürfte eine Markierung die Reproduzierbarkeit der Werte erleichtern, weshalb sich für diese Studie dafür entschieden wurde.

Durch Nummerierung war es möglich, die Punkte eines Modells in der gleichen Reihenfolge zu messen. Bei der Messung der Wirbelkörper am Patienten würde ebenfalls eine solche Reihenfolge existieren. Diese Begründung erübrigte sich im Verlauf der Datenverarbeitung, da im Nachhinein die Punkte aus statistischen Gründen als eigene Modelle gewertet wurden. Dadurch ergab sich, dass nicht alle 240 „Punktmodelle“ randomisiert sind. Die jeweils zwölf zu einem Kastenmodell zugehörigen Punktmodelle wurden stets in der gleichen Reihenfolge gemessen. Es ist möglich, dass sich dadurch systematische Fehler in die Auswertung eingeschlichen haben, bspw. durch ungeübtere oder unaufmerksamere Messungen an den jeweils ersten und letzten der zwölf Messpunkte. Dieser Einfluss sollte jedoch nicht zu hoch bewertet und nicht für die starken Schwankungen der Reliabilitäten verantwortlich gemacht werden, da stets nach bestem Gewissen aufmerksam und konzentriert gearbeitet wurde.

Das Aufsetzen der Geräte erfolgte langsam, um die Markierungen und den rechten Winkel so exakt wie möglich einzuhalten. Der nicht-messende Rater kontrollierte zu Beginn die Position des Gerätekopfes und gab gegebenenfalls Korrekturhinweise. Es wurde sich für den 90°-Winkel entschieden, da dieser leichter visuell zu kontrollieren ist, als der für den „Spineliner“ empfohlene Winkel von 45° am Patienten. Die exakte Wiederholbarkeit des Aufsatzwinkels gehört zur Standardisierung des Messablaufs, worauf in dieser Studie besonders hoher Wert gelegt wurde. Cremonese et al. 1998 nutzten in ihrer ersten Studie ebenfalls einen Aufsatzwinkel von 90°. Sie erreichten in diesem Winkel eine ausgezeichnete Reproduzierbarkeit, allerdings wurde das Gerät in eine Messapparatur gespannt. Dadurch entfiel der menschliche Anteil an der Messung. Klinisch bedeutsamer ist die Bedienung eines Geräts durch einen oder mehrere Rater.

Der Druck auf die Modelle wurde vorsichtig gesteigert, bis entweder die Vorspannung überwunden („Spineliner“) oder die Zielkraft erreicht wurde („TTM“). Dennoch ist eine unterschiedliche Taktung der Messungen zwischen Rater A und B möglich, da kein gleichmäßiger Signalton (bspw. per Metronom) genutzt wurde. Es ist denkbar, dass somit durch eine unvollständige Rückbildung der Mattenverformung des

Nachbarpunktes der zu messende Punkt beim schneller messenden Rater stärker beeinflusst wurde, als beim langsamer messenden Rater. Ein solcher Effekt nennt sich Hysteresis.

Bei den Messergebnissen wäre demzufolge eine durchgehende Raterdifferenz zu erwarten. Tatsächlich maß Rater A mit dem „TTM“ durchschnittlich niedrigere, mit dem „Spineliner“ höhere Werte als Rater B. Niedrigere Werte des „TTM“ bedeuten niedrigere Gewebespannung. Es wäre also davon auszugehen, dass Rater A in dem Falle langsamer maß und die Matten des Folgepunktes somit nicht so stark komprimiert waren. Übertragen auf den „Spineliner“ müsste Rater A dann eine niedrigere Frequenz und einen niedrigeren Widerstand messen, da diese Parameter laut Hersteller bei geringer Gewebespannung niedrige Werte zeigen. Die Messergebnisse zeigen jedoch das Gegenteil. Daher wird die Raterdifferenz durch die Hysteresis nicht, oder nur im Zusammenspiel mit weiteren Faktoren, erklärt. Weiterhin ist anzunehmen, dass jeder der Rater einmal schneller und einmal langsamer gemessen hat und sich die dadurch verursachten Fehler insgesamt ausgeglichen haben.

Dennoch kann im Hinblick auf Folgestudien auf einen gleichmäßigeren zeitlichen und größeren örtlichen Abstand zwischen den Messpunkten geachtet werden, um dem möglichen Einfluss der Hysteresis zu entgehen.

Die Ergebnisse wurden während der Messungen nicht von den Ratern eingesehen. Damit sollte verhindert werden, dass die Handhabung des Geräts durch die Messergebnisse unbewusst beeinflusst wird. Diese Verblindung wird in vielen Studien genutzt, unter anderem auch bei Leonard et al. 2003, Leach et al. 2003, Kellis et al. 2008. Kachingwe und Phillips 2005 beschäftigen sogar einen dritten Rater zur unabhängigen Verwaltung der Ergebnisse. In dieser Studie wurden die Ergebnisse nach den Messungen von Rater A in eine Datenbank übertragen. Die statistische Auswertung erfolgte nach den fünf Messterminen, um die Messungen nicht durch das Wissen über den Grad der Übereinstimmung zu beeinflussen.

Insgesamt konnte bei Aufbau und Ablauf der Studie eine hohe Standardisierung und Reproduzierbarkeit eingehalten werden.

6 Fazit und Ausblick

Die Test-Retest- und Interrater-Reliabilitäten beider Geräte konnten insgesamt nicht als ausreichend bezeichnet werden. Es gab deutliche Unterschiede zwischen den Ratern und den Messzeitpunkten. Daher ist eine, rein auf die Messwerte bezogene, Diagnosestellung nicht zu empfehlen, zumal bisher keine Normwerte ermittelt wurden.

Vielmehr sollten die Messwerte individuell und im Verlauf betrachtet werden. Die Geräte erreichten eine zufriedenstellende Intrasession-Reliabilität. Somit können sie innerhalb einer Sitzung am selben Patienten zuverlässige Ergebnisse liefern.

Beide Geräte konnten die Modellcharakteristika signifikant unterscheiden. Daher ist zu vermuten, dass die Unterschiede der Festigkeit der Wirbel bzw. Muskeln eines Patienten ebenso erkannt werden. Im Verlauf kann durch Folgemessungen der Erfolg einer Therapie beurteilt werden, bspw. durch das Angleichen eines zuvor erhöhten Messwertes in das Gesamtbild der zum selben Zeitpunkt gemessenen Werte benachbarter Wirbel oder Muskeln. Der direkte Messwertvergleich zwischen verschiedenen Sitzungen ist aus o.g. Gründen nicht empfehlenswert.

Die Beurteilung der Messwerte sollte zugunsten einer erhöhten Aussagekraft in Kombination mit der Betrachtung der Kurvendiagramme erfolgen. Diese sind ebenso nützlich, um eine bessere Visualisierung der Diagnostik und des Therapieverlaufs für den Patienten zu erreichen.

In jedem Fall bieten sich die Geräte als Diagnose- und Dokumentationsunterstützung an, können und sollen jedoch den Arzt und seine persönliche Erfahrung am Patienten nicht ersetzen.

Aus der Diskussion der Arbeit ergaben sich Anregungen für nachfolgende Studien. Eine stärkere Herausarbeitung der Modellschichten mit Zuordnung in verschiedene Gruppen kann Aussagen zur Reichweite der Messinstrumente in die Tiefe des Gewebes ergeben. Das Erstellen von Referenzwerten erscheint weniger dringlich, als das Überprüfen der Aussagekraft der Geräte im Therapieverlauf. Sowohl für weitere Arbeiten am Modell als auch Menschen sollte der Einfluss von elektromechanischen Störfaktoren bedacht werden.

7 Literatur- und Quellenverzeichnis

Aarrestad D.D, Williams M.D, Fehrer S.C, Mikhailenok E, Leonard C.T. 2004. Intra- and Interrater reliabilities of the myotonometer when assessing the spastic condition of children with cerebral palsy. Journal of child neurology 19:894-901.

Amelang M, Bartussek D, Stemmler G, Hagemann D. 2006. Differentielle Psychologie und Persönlichkeitsforschung. Stuttgart: W. Kohlhammer Verlag.

Ammer K. 2009. Die Wiederholbarkeit von Palpationsbefunden an der Wirbelsäule ist gering. Manuelle Medizin 47:349-351.

Anders H.L, Myburgh C, Hartvigsen J, Rasmussen C, Hartvig M, Marstrand K, Aagaard P. 2010. Standardized simulated palpation training - development of a palpation trainer and assessment of palpatory skills in experienced and inexperienced clinicians. Manual Therapy, 15:254-260.

Assink N, Bergman G.J.D, Knoester B, Winters J.C, Dijkstra P.U, Postema K. 2005. Interobserver reliability of neck-mobility measurement by means of the flock-of-birds electromagnetic tracking system. Journal of Manipulative and Physiological Therapeutics, 28:408-413.

Baumgartner H, Dvorak J, Graf-Baumann T, Terrier B. 1993. Grundbegriffe der manuellen Medizin. Berlin, Heidelberg: Springer Verlag.

Beyer L. 2008. Untersuchung und Behandlung mit den Händen - Varianten einer Manuellen Medizin. Manuelle Medizin, 47:73-75.

Beyer L, Tilscher H, Rizzi C, Nordmeyer V. 2008. Eigenschaften spinaler segmentaler Beweglichkeit. Manuelle Medizin, 46:105-110.

Beyer L, Nordmeyer V, Sief R, Tilscher H. 2009. Beweglichkeit der Wirbelsäulensegmente. Manuelle Medizin, 47:310-324.

Bornmann F. 2009. Klinische Ganganalyse mittels 3D-Ultraschalltopometrie, Reliabilitätsbetrachtungen und Störungsidentifikation bei SpA-Patienten [Dissertation]. Jena: Friedrich-Schiller-Universität.

Bortz J, Döring N. 2006. Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler. Vierte Aufl. Heidelberg: Springer Verlag.

Bortz J, Lienert G.A, Boehnke K. 2008. Verteilungsfreie Methoden in der Biostatistik. Dritte Aufl. Heidelberg: Springer Verlag.

Buchmann J. 2002. Manuelle Medizin und Osteopathie in Deutschland oder Was ist Neues an der Osteopathie?. Manuelle Medizin, 40:235-237.

Buchmann J, Wende K, Kundt G, Haessler F. 2005. Manual treatment effects to the upper cervical apophysial joints before, during and after endotracheal anaesthesia. American Journal of Physical Medicine & Rehabilitation, 84:251-257.

Buchmann J, Blümel G, Beyer L. 2007. Objektivierbare apparative Messung der Gewebespannung mittels Tissue Tensiometer® (TTM). Manuelle Medizin, 45:191-194.

Bühl A. 2008. SPSS 16. Elfte Aufl. München: Pearson Studium.

Bühner M. 2006. Einführung in die Test- und Fragebogenkonstruktion. Zweite Aufl. München: Pearson Studium.

The Cochrane Library. 2011. <http://www.cochrane.org>; 20.07.2011.

Conradi S, Smolenski U.C, Bak P, Biskupek H, Frey M, Galeazzi A, Klaas T, Kluge K, Müller K, Nagel J, Psczolla M, Schröder N, Sturm C, Wagner W, Wilde W. 2003. Reliabilität der manualmedizinischen Tests bei Low Back Pain (LBP)-Patienten. Manuelle Medizin, 41:272-278.

Conradi S, Smolenski U.C. 2005. Testgütekriterien manualmedizinischer Tests bei

Low-back-pain-Patienten. Manuelle Medizin, 43:227-234.

Cramer A, Doering J, Gutmann G. 1990. Geschichte der manuellen Medizin. Erste Aufl. Berlin: Springer-Verlag.

Cremonese J.G, Crunick J, Becse T, Laskey L. 1998. Examination of a piezoelectric testing device for the measurement and recording of spinal segment mobility characteristics [Pilotstudie]. Cranberry: Sigma Instruments.

Davies C. 2006. The Frozen Shoulder Workbook. Erste Aufl. Kanada: Raincoast Books.

Deutsche Gesellschaft für Manuelle Medizin (DGMM). 2011. <http://www.dgmm.de>; 22.10.2011.

Diemer F, Sutor V. 2007. Praxis der medizinischen Trainingstherapie. Erste Aufl. Stuttgart: Georg Thieme Verlag.

Dunk N.M, Chung Y.Y, Compton D.S, Callaghan J.P. The reliability of quantifying upright standing postures as a baseline diagnostic clinical tool. Journal of Manipulative and Physiological Therapeutics, 27:91-96.

Dvorak J, Gauchat M.H, Graf-Baumann T, Mühlemann D, Psczolla M. 2001. Manuelle Medizin, Chiropraktik, Osteopathie. Man Med Osteopath Med, 39:66-71.

Eder M, Tilscher H. 1990. Chirotherapie. Zweite Aufl. Stuttgart: Hippokrates Verlag.

Eliasziw M, Young S.L, Woodbury M.G, Fryday-Field K. 1994. Statistical methodology for the concurrent assessment of interrater and intrarater reliability: Using goniometric measurements as an example. Physical Therapy, 74(8):777-788.

Europäische Union. 2004. Richtlinie 2004/108/EG des europäischen Parlaments und des Rates vom 15.Dezember 2004 zur Angleichung der Rechtsvorschriften der Mitgliedstaaten über die elektromagnetische Verträglichkeit und zur Aufhebung der

Richtlinie 89/336/EWG. Amtsblatt der Europäischen Union, 390:24-37.

Fleiss J.L. 1999. The design and analysis of clinical experiments. Zweite Aufl. New York: Wiley.

Goldmann R, Bornscheuer A, Kirchner E. 1997. Gelenkblockierungen und gestörtes Gelenkspiel unter Muskelrelaxation. Manuelle Medizin, 35:56-58.

Greenmann P.E. 1998. Lehrbuch der Osteopathischen Medizin. Erste Aufl. Heidelberg: Hüthig Fachverlage.

Grimsehl E, Schallreuter W. 1991. Lehrbuch der Physik: Bd. I : Mechanik, Akustik, Wärmelehre. 27. Aufl. Leipzig: Vieweg + Teubner.

Häder M. 2010. Empirische Sozialforschung: eine Einführung. Zweite Aufl. Wiesbaden: VS Verlag.

Haerting J, Rink C, Stang A. 2007. Klinische Epidemiologie. Zweite Aufl. Bern: Verlag Hans Huber, 43-44.

Haneline M.T, Cooperstein R, Young M, Birkeland K. 2008. Spinal motion palpation: a comparison of studies that assessed intersegmental end feel vs excursion. Journal of Manipulative and Physiological Therapeutics, 31(8):616-626.

Haneline M.T, Young M. 2009. A Review of Intraexaminer and Interexaminer Reliability of Static Spinal Palpation: A Literature Synthesis. Journal of Manipulative and Physiological Therapeutics, 32(5):379-386.

Harrison D.E, Haas J.W, Harrison D.D, Holland B, Janik T. 2005. Sagittal skin contour of the cervical spine: interexaminer and intraexaminer reliability of the flexicurve instrument. Journal of Manipulative and Physiological Therapeutics, 28(7): 516-519.

Hartung J, Elpelt B, Klösener K-H. 2005. Statistik. 14. Aufl. Oldenbourg: Wissenschaftsverlag.

Heimann D. 2008. Leitfaden Manuelle Medizin. Vierte Aufl. München: Elsevier, Urban & Fischer Verlag.

Horstkotte J. 2011. <http://www.ce-zeichen.de>.

Internationale Gesellschaft für Elektrosmog-Forschung. 2011. <http://www.elektrosmog.com/de>.

Jüngling N, Smolenski U.C, Loth D. Untersuchung zur Reliabilität und Validität der dreidimensionalen Funktionsanalyse des Kiefergelenks. Manuelle Medizin, 42:441-448.

Kachingwe A.F, Phillips B.J. 2005. Inter- and intrarater reliability of a back range of motion instrument. Archives of Physical Medicine and Rehabilitation, 86:2347-2353.

Kayser R. 2010. Manuelle Medizin. Orthopäde, 39(2):217-228.

Kellis E, Adamou G, Tziliou G, Emmanouilidou M. 2008. Reliability of spinal range of motion in healthy boys using a skin-surface device. Journal of Manipulative and Physiological Therapeutics, 31:570-576.

Krämer J. 2006. Bandscheibenbedingte Erkrankungen: Ursachen, Diagnose, Behandlung, Vorbeugung, Begutachtung. Fünfte Aufl. Stuttgart: Georg Thieme Verlag.

Kumar K. 1996. Historical Perspective Spinal Deformity and Axial Traction. Spine, 21(5):643-655.

Lantz C.A, Klein G, Chen J, Mannion A, Solinger A.B, Dvorak J. 2003. A reassessment of normal cervical range of motion. Spine, 28(12):1249-1257.

Leach R.A. 2002. Differential compliance instrument in the treatment of infantile colic: a report of two cases. Journal of Manipulative and Physiological Therapeutics,

25(1):58-62.

Leach R.A, Parker P.L, Veal P.S. 2003. PulStar differential compliance spinal instrument: a randomized interexaminer and intraexaminer reliability study. *Journal of Manipulative and Physiological Therapeutics*, 26(8):493-501.

Leonard C.T, Deshner W.P, Romo J.W, Suoja E.S, Fehrer S.C, Mikhailenok E.L. 2003. Myotonometer Intra- and Interrater Reliabilities. *Archives of Physical Medicine and Rehabilitation*, 84:928-932.

Lewit K. 1997. *Manuelle Medizin*. Siebte Aufl. Heidelberg/Leipzig: Johann Ambrosius Barth Verlag.

Lidström Å, Ahlsten G, Hirschfeld H, Norrlin S. 2009. Intrarater and Interrater reliability of myotonometer measurements of muscle tone in children. *Journal of child neurology* 24(3):267-274.

Lienert G.A, Raatz U. 1998. *Testaufbau und Testanalyse*. Sechste Aufl. Weinheim: Psychologie Verlags Union.

Meyers C.R, Blesh T.E. 1962. *Measurement in physical education*. Erste Aufl. New York: Ronald Pr.

Muscolino J.E. 2009. *The Muscle and Bone Palpation Manual – with Trigger Points, Referral Patterns, and Stretching*. Erste Aufl. St.Louis/ USA: Mosby, Elsevier Verlag.

Najm W.I, Seffinger M.A, Mishra S.I, Dickerson V.M, Adams A, Reinsch S, Murphy L.S, Goodman A.F. 2003. Content validity of manual spinal palpatory exams - a systematic review. *BMC Complementary and alternative medicine*, Mai 2003.

Nicholson L.L, Adams R.D, Maher C.G. 2003. Manual discrimination capability when only viscosity is varied in viscoelastic stiffness stimuli. *Journal of Manipulative and Physiological Therapeutics*, 26:365-373.

NTI. 2011. <http://www.neurogenic.com>.

Oschmann K. 2010. Untersuchung der Beweglichkeit der oberen Extremität mittels 3D-Bewegungsfunktionsanalysesystem Zebris® und Winkelmessung [Dissertation]. Jena: Friedrich-Schiller-Universität.

Patijn J. 2002. Studien zur Reproduzierbarkeit und Validität diagnostischer Verfahren in der Manuellen Medizin. Manuelle Medizin, 40:339-351.

PONS. 2008. <http://www.pons.eu>.

Portney L.G, Watkins M.P. 2000. Foundations of clinical research: applications to practise. Michigan: Prentice Hall Health.

ProAdjuster Lifestyle. 2011. <http://www.proadjusterlifestyle.com/index.php>. 02.10.2011.

ProAdjuster University. 2011. <http://www.proadjusteruniversity.com/>. 02.10.2011.

Psczolla M. 1997. Die Versorgungssituation der Manuellen Medizin und Manuellen Therapie in Deutschland. In: Dvorak J, Hrsg. Manuelle Medizin. Dritte Aufl. Stuttgart: Georg Thieme Verlag, 230-237.

Pubmed. 2011. <http://www.ncbi.nlm.nih.gov/pubmed>.

PulStar. 2011. <http://www.pulstar.us>.

Ronchi A.J, Lech M, Taylor N.F, Cosic I. 2008. A reliability study of the new back strain monitor based on clinical trials. Conference Proceedings - IEEE Engineering in Medicine and Biology Society, 2008:693-696.

Rustler T. 2010. Willkommen zum Spineliner® -Seminar! [Broschüre]. Wien: TTC GmbH.

Rustler T, Tilscher H. 2004. Treatment of the cervical spine with the Spineliner [Pilotstudie]. Wien: Ludwig Boltzmann Institut. Präsentiert: FIMM-Kongress 2004, Bratislava.

Schmitt W.H, Cuthbert S.C. 2008. Common errors and clinical guidelines for manual muscle testing: "the arm test" and other inaccurate procedures. Chiropractic and Osteopathy, Dezember 2008.

Selbmann H.-K, Meisner C. 2000. Wie beurteilt man die Qualität klinisch-therapeutischer Studien? Internist, 41:344-348.

Shrout P.E, Fleiss J.L. 1979. Intraclass correlations: Uses in assessing rater reliability. Psychological Bulletin, 86:420-428.

Sigma Instruments. 2011. <http://www.sigma-instruments.com/index.php>.

Simons D.G, Travell J.G, Simons L.S. 1999. Myofascial Pain and Dysfunction. USA: Williams & Wilkins.

Simons D.G, Travell J.G, Simons L.S. 2002. Handbuch der Muskel-Triggerpunkte. Obere Extremität, Kopf und Rumpf. München: Urban & Fischer Verlag/ Elsevier GmbH.

Simons D.G, Travell J.G. 2004. Flipchart Triggerpunkte: Muskel-Triggerpunkte und ihre Schmerzfelder. München: Urban & Fischer Verlag/ Elsevier GmbH.

Smolenski U.C, Endres G, Bocker B. Untersuchung der Halswirbelsäulenbeweglichkeit mittels Bewegungsfunktionsanalysesystem *zebris* und Winkelmessung. Manuelle Medizin, 41:365-373.

Speckmann E.-J, Wittkowski W. Das Substrat der „Blockierung“. Manuelle Medizin, 35:176-183.

Squires G.L. 1971. Messergebnisse und ihre Auswertung: eine Anleitung zum

praktischen naturwissenschaftlichen Arbeiten. Erste Aufl. Berlin: Walter de Gruyter.

Steinlechner K. 1938. Myotonometer – Ein neuer Apparat zur Messung des Muskeltonus. *Experimental Medicine*, 103/3:400-404.

Stochkendahl M.J, Christensen H.W, Hartvigsen J, Vach W, Haas M, Hestbaek L, Adams E, Bronfort G, Beyer L. 2007. Manuelle Untersuchung der Wirbelsäule. *Manuelle Medizin*, 45:301-308.

Taylor W.R, Consmüller T, Rohlmann A. 2010. A novel system for the dynamic assessment of back shape. *Medical Engineering & Physics*, 32:1080-1083.

Tepe R, Zhang J. A pilot investigation of expert and novice intraexaminer and interexaminer reliability of durometer analysis of the cervical spine. Chesterfield: Logan college of chiropractic.

TTC. 2008. <http://www.spineliner.com>.

Vogel T. 2007. Untersuchungen zur Reliabilität und Validität der Gleichgewichtsanalyse mittels ultraschalltopometrischer Cranio-Corpo-Graphie (zebris®) [Dissertation]. Jena: Friedrich-Schiller-Universität.

Web of Knowledge. 2011. <http://wokinfo.com>.

Will T. 2004. Evaluierung der Ultraschall-3D-Topometrie (zebris®) im Vergleich zur Funktionsradiographie der Lendenwirbelsäule [Dissertation]. Jena: Friedrich-Schiller-Universität.

Wirtz M, Caspar F. 2002. Beurteilerübereinstimmung und Beurteilerreliabilität. Erste Aufl. Göttingen: Hogrefe-Verlag.

Zahornitzky W, Keusch R, Tilscher H, Beyer L. 2010. Behandlung des oberen Zervikalsyndroms mit funktioneller computerassistierter Reflextherapie (Spineliner). *Manuelle Medizin*, 48:31-38.

8. Anhang

Ehrenwörtliche Erklärung

Hiermit erkläre ich, dass mir die Promotionsordnung der Medizinischen Fakultät der Friedrich-Schiller-Universität bekannt ist,

ich die Dissertation selbst angefertigt habe und alle von mir benutzten Hilfsmittel, persönlichen Mitteilungen und Quellen in meiner Arbeit angegeben sind,

mich folgende Personen bei der Auswahl und Auswertung des Materials sowie bei der Herstellung des Manuskripts unterstützt haben: Sebastian Ullrich, Prof. Dr. med. Ulrich Smolenski, Prof. Dr. med. Lothar Beyer, Kristin Neumer, Dr. Mario Walther,

die Hilfe eines Promotionsberaters nicht in Anspruch genommen wurde und dass Dritte weder unmittelbar noch mittelbar geldwerte Leistungen von mir für Arbeiten erhalten haben, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen,

dass ich die Dissertation noch nicht als Prüfungsarbeit für eine staatliche oder andere wissenschaftliche Prüfung eingereicht habe und

dass ich die gleiche, eine in wesentlichen Teilen ähnliche oder eine andere Abhandlung nicht bei einer anderen Hochschule als Dissertation eingereicht habe.

Jena, 04.09.2012

Anne-Maria Zschiegner